Vol. 53 JUNE 1962 No. 3

Journal of Educational Psychology

CONTENTS

Spelling Errors and the Serial-Position Effect 105
ARTHUR R. JENSEN

The Effects of Response Mode and Response Difficulty on Programed Learning 110
ROBERT A. GOLDBECK AND VINCENT N. CAMPBELL

Concept Attainment by Deaf and Hearing Adolescents 119
SOLIS L. KATES, LEE YUDIN, AND RONALD K. TIFFANY

A Comparison of Two Response Modes in an Autoinstructional Program with Children in the Primary Grades 127
EVAN R. KEISLAR AND JOHN D. MCNEIL

The Prediction of the Academic, Artistic, Scientific, and Social Achievement of Undergraduates of Superior Scholastic Aptitude 132
JOHN L. HOLLAND AND ALEXANDER W. ASTIN

Some Non-Intellectual Correlates of Academic Achievement Among Mexican-American Secondary School Students 144
LOIS J. GILL AND BERNARD SPILKA

The Effects of Anxiety, Achievement Motivation, and Task Importance upon Performance on an Intelligence Test 150
LAWRENCE S. WRIGHTSMAN, JR.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

JOHN B. CARROLL
Harvard University
LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
DOROTHEA MCCARTHY
Fordham University
CARSON MCGUIRE
University of Texas
PAUL H. MUSSEN
University of California, Berkeley
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University
BENTON J. UNDERWOOD
Northwestern University

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association.*

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Acting Production Manager

SONDRA SEGALL
Editorial Assistant

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

P. O. Box 1064, Richmond 8, Va.
and 1333 Sixteenth Street, N.W., Washington 6, D. C.
Second-Class postage paid at Richmond, Va.

Copyright, 1962 by the American Psychological Association, Inc.

JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 53 June 1962 Number 3

SPELLING ERRORS AND THE SERIAL-POSITION EFFECT[1]

ARTHUR R. JENSEN
University of California

The relative frequencies of spelling errors as a function of letter position have been examined for 7-, 9-, and 11-letter words selected at random from the Thorndike-Lorge word list. These were administered to 150 8th graders, 158 10th graders, and 89 junior college freshmen, respectively. The distribution of errors according to letter position was found to closely approximate the classical, skewed, bow-shaped, serial-position curve for errors generally found in serial rote learning. Other features in common between spelling and serial learning were discussed. It is suggested that a theory of serial learning and of the serial-position effect may be germane to the psychology of spelling.

Perusal of the research on spelling reveals that this subject has been traditionally dominated more by a linguistic than by a psychological orientation (Horn, 1960). Spelling difficulties have been analyzed largely in terms of phonetics, syllabication, parts of speech, and frequency of word usage. Spelling research is also predominantly characterized by a rather simple, unanalytical, correlational approach. Relationships have been found between measures of spelling ability and measures of reading, MA, IQ, achievement in various school subjects, and numerous other variables (Horn, 1960). Seldom, however, is spelling studied as a learning phenomenon or examined in terms of principles derived from the experimental psychology of learning. The investigator has conceived of what might be called a psychological theory of spelling based primarily on concepts in learning theory. The purpose of the present investigation is to examine only one aspect of the formulation, viz., the serial-position effect. The similarity of the distribution of spelling errors, when errors are plotted according to their position in the word, to the distribution of errors in serial rote learning—the so-called serial-position effect—has never before been noted. The correspondence between spelling and serial learning would seem to offer some clues concerning the nature of spelling errors.

When a series of words, nonsense syllables, symbols, or figures is learned, the errors made before mastery are distributed according to the position of each item in the series, producing the familiar, skewed, bow-shaped serial-position curve. Various theories have been proposed to account for this phenomenon (Jensen, 1962a; McGeoch & Irion, 1952, pp. 125-134). Since a word consists of a series of letters, and since spelling a word consists of putting each let-

[1] This study was supported in part by a Faculty Research Grant from the University of California and in part by the National Science Foundation.

ter in its proper order, we might ask whether the serial position effect, which is generally manifested in every form of serial learning investigated in the laboratory, is also manifested in spelling. If so, the distribution of spelling errors should be similar to the serial-position curves produced by the serial learning of other materials. That is to say, errors should generally be most frequent in the middle of a word, with fewest errors at the beginning and end of the word. That spelling errors occur more frequently in the middle of words has probably been noted before, but apparently not with sufficient precision to reveal the more subtle features of the serial-position effect.

Variations in the shape of the serial-position curve are mainly a function of the number of items in the series. The curve for a short series is markedly skewed, with the largest proportion of errors occurring well past the middle of the series. As the series becomes longer, the serial-position curve becomes more symmetrical (McGeoch & Irion, 1952, p. 123). It would be interesting indeed if this were found to be the case also for the distribution of spelling errors in words of different lengths. The present investigation was addressed to this question.

METHOD

Spelling

So that word selection would be unbiased with respect to the purpose of the present study, every 7-letter, 9-letter, and 11-letter word of the Thorndike-Lorge word list (1944, pp. 1-208) was typed on a slip of paper. The slips were thoroughly shuffled to eliminate alphabetical order, and then 75 eight-item word lists were made up separately for 7-, 9-, and 11-letter words by drawing the words at random from the total word pool. Foreign words and proper nouns were eliminated.

In order to maximize spelling errors while at the same time presenting words that were reasonably familiar to the subjects, it was decided to give the 7-letter words only to the eighth graders ($N = 150$), the 9-letter words only to tenth graders ($N = 148$), and the 11-letter words only to junior college freshmen ($N = 89$).

The spelling tests were selected at random from all the eight-item spelling lists. Every eighth grader was given 24 7-letter words; every tenth grader was given 24 9-letter words, but different lists were used in different classes, so the total number of different 9-letter words was 96. Each of the junior college students was given 48 11-letter words.

The subjects had mimeographed forms on which to print the words. For a 7-letter word, for example, a row of seven adjacent squares was provided, with the instructions to put one letter in each square. Similar forms were used for 9- and 11-letter words. This procedure permitted the position of each spelling error to be easily tabulated.

For all groups of subjects, the words were read aloud by the regular classroom teacher in the fashion of the usual spelling test. Students were not given the words for study before the test. Neither the teachers nor the students knew the purpose of the test.

Serial Learning

To permit a detailed comparison between the position of errors in spelling and in serial rote learning, serial learning data were obtained. Sixty subjects (college students) learned by the anticipation method a nine-item serial list composed of nine colored geometric forms: triangles, squares, and circles, colored red, blue, and yellow. The stimuli were projected automatically one at a time at a 3-second rate with a 6-second intertrial interval. (A detailed description of the apparatus is given elsewhere—Jensen, Collins, & Vreeland, 1962.) Never were figures of the same shape or color adjacent in the series. A green light of 3-seconds duration served as the signal for the subject to anticipate the first item in the series. Subjects responded by saying "red triangle, blue square," etc. All subjects learned to a criterion of one perfect trial. Two different orders of stimuli were used, with 30 subjects assigned to each order.

RESULTS

All the serial-position curves are presented as the percentage of total errors that occur at each serial position, so that the area under all the curves is the

same. This method of presentation was adopted since the interest here does not concern the absolute number of errors but only the relative shapes of the curves for serial learning errors and for spelling errors. For the color-form serial-learning data, the percentage of errors at each position was determined for each subject individually and then averaged over all subjects, in order to weight each subject equally in the group curve. A similar method was applied to the spelling data. Since some words were misspelled much more frequently than others, the percentage of errors at each position was determined for each word and these were averaged over all words so that every word is weighted equally in the final curve.

Figure 1 shows the distribution of errors as a function of letter position for 96 9-letter words and for the serial position of errors made by 60 subjects in learning the color-form series. Both curves evince the essential features of the serial-position effect, that is, a bowed curve with the greater proportion of errors occurring after the middle position. The intraclass correlation between the pairs of data points for the two curves may be regarded as an index of the degree of similarity in the shapes of the curves (Haggard, 1958). The intraclass r was .81 between the spelling and the color-form curves. When the curves produced by the two serial orders of the color-form test were intercorrelated, the r was .96, which may be regarded as an equivalent-forms reliability of the shape of the serial-position curve.

The error curves for 7- and 11-letter words are shown in Figure 2. These curves closely resemble serial-position curves found in the literature on serial learning (Jensen, 1962a; McGeoch & Irion, 1952, p. 123) with the degree of skewness varying as a function of the number of items in the series. The peak of errors for a 7-item list is generally at

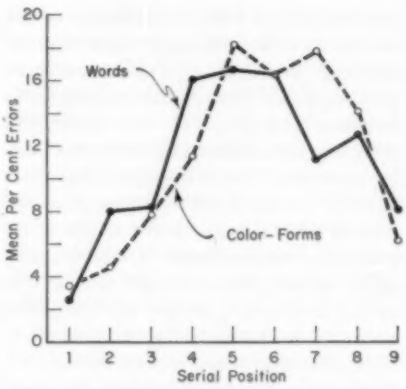

FIG. 1. Serial-position curves for errors in nine-item color-form serial learning (60 subjects) and for spelling errors in nine-letter words. (The latter is based on a total of 3,864 spelling errors, i.e., incorrect letters, made on 96 words by 148 tenth graders.)

Position 5, and for an 11-item list it is at Position 7. The spelling errors conform perfectly to these expectations.

Since the usual method of learning to spell words is not the anticipation method generally used in serial learning experiments, but involves simultaneous viewing of all the letters in the word, one might ask if learning the color-forms by this method also evinces the serial-position effect. If it did not, it would

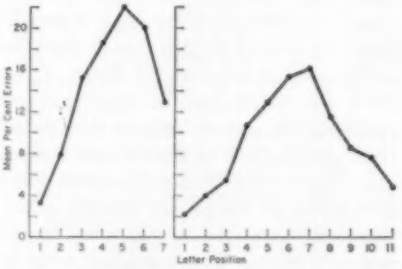

FIG. 2. Serial-position curves for errors in 7- and 11-letter words. (The curve for 7-letter words is based on a total of 1,546 spelling errors made on 24 words by 150 eighth graders. The 11-letter curve is based on a total of 2,262 spelling errors made on 48 words by 89 junior college freshmen.)

be doubtful that the resemblance of the error curves for spelling to those of serial learning was indicative of a common psychological process underlying both forms of learning. To investigate this question, an additional 30 subjects (college students) were given a color-form test that involved simultaneous presentation of all the items in the series. Nine colored plastic forms (red triangle, green square, blue triangle, yellow diamond, red circle, yellow triangle, blue circle, pink square, and green diamond) were placed in a row before the subject. After the subject had studied the order of the objects for 10 seconds, the objects were disarranged haphazardly and the subject was asked to rearrange them in the proper order. This procedure was repeated, always with the same serial arrangement of the objects, until the subject learned to reconstruct their order perfectly. There was a different order for each subject.

Figure 3 shows the serial-position curve obtained by this method of simultaneous presentation. It may be regarded as a spatial, rather than the usual temporal, form of serial learning. The serial-position effect clearly is not limited to learning by the anticipation method (Jensen, 1962b). It should be noted that the curve in Figure 3 more closely resembles the curve for misspellings in words in Figure 1 than does the serial anticipation curve in Figure 1. In brief, the distribution of spelling errors more closely corresponds to the spatial serial-position effect than to the temporal. This is what we would expect, since the letters of words are viewed simultaneously as in the spatial serial learning procedure. The intraclass correlation between the words curve in Figure 1 and the curve in Figure 3 is .90, which indicates a degree of correspondence between these curves that is almost as high as the "reliability" of the curves would permit.

DISCUSSION

A strictly linguistic, rather than psychological, interpretation of these findings is not readily apparent. The obvious linguistic explanation of these results is in terms of the relative easiness of spelling the common prefixes and suffixes in words, which would tend to make for fewer spelling errors at the beginnings and ends of words having these elements. This explanation is unconvincing for several reasons. If prefixes and suffixes, most of which are composed of two or three letters, are learned as units, and if these prefixes and suffixes account for the shape of the error curve, then we should expect quite a differently shaped curve than was actually found in the present study. Each of the first three and last three letter positions should have an almost uniformly small proportion of errors. Furthermore, the differences in the skewness of curves for words of different lengths would be difficult to explain by the prefix-suffix hypothesis. It is apparent that even if the first and last three positions of the error curves are ignored, the middle portion of the word still manifests the bowed serial-position

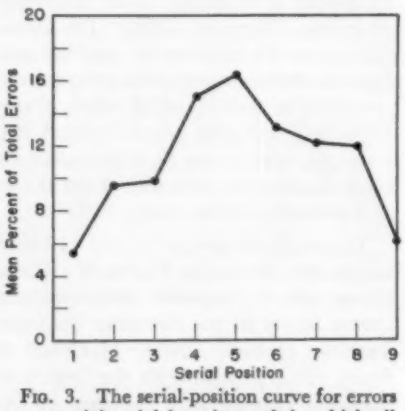

FIG. 3. The serial-position curve for errors on a spatial serial learning task in which all nine items were presented simultaneously on each trial (30 subjects).

effect. Note the curve for 11-letter words in Figure 2; even without the first and last three positions the errors are distributed in a fashion typical of serial-position curves. Furthermore, only a small proportion of the words used in this study had either suffixes or prefixes. Using as a criterion a list of 46 common Greek and Latin prefixes (McCrimmon, 1950, p. 186), it was found that only 16% of the 9-letter words had prefixes. The mean number of errors on the first three positions for these words was 5.60; for nonprefix words the mean number of errors was 6.80. This small difference, in addition to the fact that only 16% of the words had prefixes, could not account for the markedly bow-shaped error curves found in the present study.

It is not claimed that spelling errors are solely a function of serial position. Certainly some words are phonetically more difficult than others, and this factor is undoubtedly a large source of spelling errors. Each word has its own peculiar difficulties, and it appears that these difficulties can occur at any letter position in the word. It is only when the errors that are peculiar to individual words are "averaged out" by combining the position errors of a large number of words, as was done in the present experiment, that the distribution of errors associated with the serial-position effect becomes evident. Thus, a phonetically difficult element is probably more or less difficult according to its position in the word.

There is still another interesting point of correspondence between spelling and serial learning. A number of studies (reviewed by Horn, 1960, pp. 1346-1347) have found that emphasizing difficult parts of words, in presenting the words in a lesson, by underlining, writing the letters in capitals, using boldfaced type, etc., is of little or no value in increasing the probability that the word will be spelled correctly. Emphasis of particular letters in a word corresponds to the so-called von Restorff procedure in serial learning, in which an item in the series is perceptually emphasized in some manner. The emphasized item is usually learned more readily than would be a nonemphasized item in the same position, but the difficulty of learning the list as a whole remains the same as if no item had been emphasized (Jensen, 1962c; Newman & Saltz, 1958). When errors are reduced at one position, they are correspondingly increased at other positions, so that there is no overall advantage from emphasizing a particular element in the series. This seems to be true in spelling as well as in serial learning.

REFERENCES

HAGGARD, E. A. *Intraclass correlation and the analysis of variance.* New York: Dryden, 1958.

HORN, E. S. In C. W. Harris (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1960. Pp. 1337-1354.

JENSEN, A. R. An empirical theory of the serial-position effect. *J. Psychol.*, 1962, **53**, 127-142. (a)

JENSEN, A. R. Temporal and spatial serial-position effects. *Amer. J. Psychol.*, 1962, in press. (b)

JENSEN, A. R. The von Restorff isolation effect with minimal response learning. *J. exp. Psychol.*, 1962, in press. (c)

JENSEN, A. R., COLLINS, C. C., & VREELAND, R. W. A multiple S-R human learning apparatus. *Amer. J. Psychol.*, 1962, in press.

MCCRIMMON, J. M. *Writing with a purpose.* Boston: Houghton Mifflin, 1950.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

NEWMAN, S. E., & SALTZ, E. Isolation effects: Stimulus and response generalization as explanatory concepts. *J. exp. Psychol.*, 1958, **55**, 467-472.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received October 30, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 3, 110-118

THE EFFECTS OF RESPONSE MODE AND RESPONSE DIFFICULTY ON PROGRAMED LEARNING[1]

ROBERT A. GOLDBECK AND VINCENT N. CAMPBELL
American Institute for Research

2 experiments in junior high school classes compared overt, covert, and reading response modes. In Experiment I, 63 Ss completed a program of independent facts at 1 of 3 levels of difficulty. An analysis of covariance of test scores showed an interaction between response mode and difficulty ($p < .05$), the overt group performing below the other groups at the low difficulty level and above the other groups at the intermediate difficulty level. In Experiment II, 62 Ss completed a continuous discourse program on light. The 3 response modes plus a 4th option mode were used in a simple randomized design. The reading group surpassed the other groups on both immediate (NS) and 10-week retention ($p < .05$) tests. In both experiments the reading groups learned most per time spent ($p < .01$).

One of the major advantages claimed for automated teaching is that it provokes the student to respond actively to the material to be learned. The characteristic of automated techniques that accomplishes this goal is the requirement for the student to make some response during learning. In this regard, it is typically intended to prepare the learning material such that all students will be able to respond correctly nearly all of the time (Skinner, 1958).

There has been some question of the simple rationale that the most effective learning will be produced by highly cued and prompted material requiring overt responses which will be predominantly correct (Briggs, 1960; Goldbeck & Briggs, 1960; Pressey, 1960). Evans, Glaser, and Homme (1960) and Silverman and Alter (1960) have obtained data which fail to show the superiority of using overt responses in a simple learning situation. Angell and Lumsdaine (1960) have shown that prompting on three out of four trials is superior to prompting on all trials, following Cook's work (Cook, 1958; Cook & Kendler, 1956; Cook & Spitzer, 1960) showing that prompting on three out of four trials is superior to no prompting.

The experiments to be reported here concern the effects of response mode on programed learning. The overt written response mode will be compared with several other variants. The option response mode is most clearly related to the overt response mode. In the option mode the subject may write his response in the blank if he is sure it is correct, or if not sure he may leave the blank empty and proceed directly to view the feedback. A second variant is the covert response mode. Here the subject simply attempts to "think" of the correct response before viewing the feedback. The third variant is much like conventional reading in that the correct response is printed in the blank. No feedback as such is provided.

In Experiment I the overt, covert, and reading response modes are exam-

[1] This investigation was conducted by the American Institute for Research under Contract Nonr-3077(00) with the Office of Naval Research, Leslie J. Briggs, principal investigator. The investigation of the option response mode was part of a research project supported by Grant 736103.00 of the United States Office of Education, Department of Health, Education, and Welfare.

ined using independent facts as the learning content. Since each fact can be learned independently of the others, this kind of material lends itself well to variation in response difficulty by varying the amount of cueing and prompting used. Experiment I is designed to test the hypothesis that the efficacy of a response mode is dependent on the cueing characteristics that determine learning response difficulty.

This interdependency would seem to be most likely for the overt response mode. For highly cued or prompted responses, there is little chance of an error response occurring. In fact, the overt response may be performed in quite a perfunctory manner, with little or no "active" responding other than the motor response of writing. Such responses may be considered copied responses rather than constructed responses and may produce an "illusion of learning." With responses that are not so heavily cued, it is to be expected that constructing a correct overt response will produce a more relevant type of active participation, and at the same time, more errors during learning. If the more relevant response pattern facilitates learning to a greater extent than error responses degrade learning, some intermediate amount of cueing or prompting should be superior to highly cued responses. As cues and prompts are withdrawn further, error responses may override the advantages of the active overt response and learning should fall off.

In Experiment II, the overt, option, covert, and reading response modes are examined using continuous discourse material in order to extend the basic response mode findings of Experiment I to a more generally used type of learning material. The notion about response difficulty discussed above led us to construct a form of continuous discourse programing with greater response difficulty than is typical. The option response mode is intended to increase the time efficiency of overt responding and to mitigate any adverse effect of error responses by allowing the subject to omit especially difficult responses.

EXPERIMENT I

Method

Material. The learning material consisted of 35 independent factual items of information covering men of history, geography, etc. They were presented in booklet form, one fact to a page, and in the same sequence for all subjects. For each response mode (overt, covert, and reading) each fact was prepared in easy, intermediate, and difficult versions. The difficult frames (pages) provided minimum cueing. Additional cues and prompts were added to obtain the intermediate and easy level frames.

Three versions of one fact are shown below as examples of the three difficulty levels:

A major industrial center for the country, ... is the second largest city in Mexico. (Difficult)

The Spanish settlers in Mexico named this city after a province and its capital city in Spain. A major industrial center for the country, ...is the second largest city in Mexico. (Intermediate)

The ideal climate of Guadalajara has made it a popular resort area for Americans. The Spanish settlers in Mexico named this city after a province and its capital city in Spain. A major industrial center for the country, ... is the second largest city in Mexico. (Easy)

In the overt and covert response booklets, each frame was followed by an answer page showing the correct word for the blank in that frame. In the reading response booklet, the critical word was printed in the context of the frame as in conventional reading matter, with the word underlined. No answer page followed the reading response frame. The three response mode conditions, each in combination with the three difficulty levels, resulted in nine different kinds of learning booklets.

The criterion of amount learned was a test consisting of one question for each of the 35 facts. The correct answer to the question was

the response word from the corresponding learning frame. The question used for the frames given as an example above was: "What is the second largest city in Mexico?"

Subjects. A total of 63 subjects from two seventh grade classes participated in the study. Both the classes were of above-average ability, as shown by mean grade level scores of 8.45 and 8.49 for the two classes on the California Reading Maturity Test.

Procedure and design. The experiment was conducted concurrently in two separate classrooms. Subjects were assigned randomly among the nine experimental conditions, six subjects to each cell. The remaining nine subjects were assigned to the overt response mode conditions as part of a separate analysis of overt responses. Subjects worked through the learning booklet at their own speed and could not return to any page once having completed it. As soon as the subject had completed the learning booklet, the experimenter collected it and recorded the time. The subject completed the test at his own speed and was allowed to return to any question which he could not answer.

Results

Learning responses. Error rate during learning differed sharply among difficulty levels in the overt response mode, as planned. Mean numbers of correct learning responses (where 35 represents a perfect score) for the difficult, intermediate, and easy learning material groups were 6.7, 14.6, and 27.7, respectively ($F = 15.03$, $p < .01$, $df = 2/24$).

Test scores. Scores on the California Reading Maturity Test were used in a covariance analysis of criterion test scores. The average within-group correlation between the reading scores and test scores was .55. Adjusted mean test scores are shown in Table 1. The only significant effect was the Response Mode × Difficulty Level interaction ($F = 4.25$, $p < .01$, $df = 4/35$). The essence of this interaction is that the overt response group is superior at the intermediate level of difficulty and inferior at the easy level.

TABLE 1

MEAN TEST SCORES ADJUSTED BY THE CALIFORNIA READING MATURITY TEST: EXPERIMENT I

Response mode	Difficulty level		
	Easy	Intermediate	Difficult
Overt	12.5	25.5	20.3
Covert	20.9	17.1	21.3
Reading	20.6	21.5	22.8

Note.—Estimated *SD* within cells = 4.7.

At the low difficulty level, the overt response group mean is significantly lower than means for either of the other response mode groups. Using the covariance data with $df = 35$ to test adjusted mean differences (Lindquist, 1953), $t = 2.71$, $p < .05$ and $t = 2.85$, $p < .01$ for overt group differences from the covert group and reading group, respectively. Requiring constructed responses for the learning material for which constructing a response was relatively easy resulted in significantly poorer learning as measured by an immediate retention test.

As constructing a response became more difficult, at the intermediate difficulty level, the test performance of the overt response group improved dramatically and significantly ($t = 4.45$, $p < .01$). At this level of difficulty, the overt response group performed significantly better than the covert group ($t = 2.92$, $p < .01$). The value of requiring a constructed response apparently depends on the conditions under which a response is made.

Test response—learning response relationship. In order to evaluate the effect of correct versus incorrect learning responses on test performance, an analysis was performed on the learning and test responses for all 27 subjects who were required to give overt learning responses. At each difficulty level, a

TABLE 2

PERCENTAGE CORRECT OF TEST RESPONSES PRECEDED BY CORRECT VERSUS INCORRECT LEARNING RESPONSES: EXPERIMENT I

Test responses	Difficulty level		
	Easy	Intermediate	Difficult
Preceded by correct learning responses	54	74	85
Preceded by incorrect learning responses	36	63	49

total of 315 learning and test response pairs are available. These pairs were dichotomized according to whether the learning response was correct or incorrect. The number of correct test responses following a correct learning response and the number following an incorrect learning response were determined, and were expressed as percentages. These data are presented in Table 2.

While correct learning responses result in a higher percentage of correct test responses at each difficulty level, it is apparent that this effect was outweighed by other factors affecting test performance. In fact, test performance of the intermediate level group following *incorrect* learning responses was higher than test performance of the easy level group following *correct* learning responses (see Table 2).

Learning time. The summary of learning time data is shown in Table 3. Note that the learning time means for the overt group follow the same pattern as the adjusted test score means. As might be expected, due to the necessity to think of a response and to write it out, the overt response groups spent more time in learning than the other groups. It is interesting to note that the covert groups spent no more time on learning than the reading groups.

TABLE 3

MEAN LEARNING TIMES (IN MINUTES): EXPERIMENT I

Response Mode	Difficulty level		
	Easy	Intermediate	Difficult
Overt	22.3	31.0	19.2
Covert	16.5	15.3	14.3
Reading	17.3	16.0	15.2

Note.—Estimated *SD* within cells = 4.9.

TABLE 4

ANALYSIS OF VARIANCE SUMMARY FOR LEARNING TIME DATA: EXPERIMENT I

Source	*df*	*MS*	*F*
Response mode	2	425.0	17.63**
Difficulty level	2	93.7	3.89*
Class	1	31.2	1.29
R × D	4	72.8	3.02*
R × C	2	60.0	2.49
D × C	2	8.5	
R × D × C	4	19.0	
Within	36	24.1	

*$p < .05$.
**$p < .01$.

Table 4 presents the analysis of variance for these data, showing a significant Response Mode × Difficulty Level interaction.

Learning efficiency. Learning efficiency scores were obtained by dividing the test score for each subject by his learning time score, giving a measure of amount learned per unit of learning time. Mean efficiency scores are shown in Table 5. An analysis of variance indicated a clear-cut distinction among the three response modes ($F = 13.74$, $p < .01$, $df = 2/36$), with the reading mode most efficient, the overt response mode least efficient, and the covert response mode falling between.

TABLE 5

MEAN LEARNING EFFICIENCY SCORES (TEST SCORE/LEARNING TIME): EXPERIMENT I

Response Mode	Difficulty level		
	Easy	Intermediate	Difficult
Overt	0.67	0.82	1.12
Covert	1.19	1.15	1.32
Reading	1.37	1.60	1.60

Note.—Estimated *SD* within cells = 0.37.

Test time. Inasmuch as no time limits were imposed on test performance, it was possible to determine whether the experimental conditions during learning had any effect on the time required to complete the test. A 3 × 3 analysis of variance on test score times for the response mode and difficulty variables showed a significant difference only among response modes ($F = 4.74$, $p < .05$, $df = 2/45$). The means for the overt, covert, and reading response mode conditions were 10.4 minutes, 13.4 minutes, and 13.6 minutes, respectively.

EXPERIMENT II

Method

Material. The learning material consisted of 32 frames on the topic of light, based on the initial portion of an autoinstructional program developed by Klaus, Holmok, Lumsdaine, Meyer, & Quisenberry (1960). The modified program, in which a frame had from one to four response terms, contained a total of 51 response terms.

The learning booklets followed the same format as in Experiment I, except for the feedback terms. The correct words for the blanks were printed on the right side of the page, in line with the appropriate blank, and were covered by a slip of gray paper. This cover could be lifted by the subject to show the correct words. As in Experiment I, there were no feedback terms for the reading response group.

The criterion test of amount learned was a mixture of recall and recognition items. Although there was not a one-to-one correspondence between test items and learning frames, the test was designed to assign score values homogeneously among the areas of learning content. Some of the test answers had been a "response" term in the learning material while others had not.

Subjects. Data were collected from a total of 62 subjects in two eighth grade classes. The ability level of the two classes was comparable ($t = .75$) as measured by mean score on the reading portion of the School and College Ability Test (SCAT). Both classes were above average on this test.

Procedure and design. The experiment was conducted concurrently in the two separate classrooms. Subjects were assigned randomly to the four experimental conditions: overt, option, covert, and reading response modes. The order of events for all subjects was (*a*) a practice program and test on mountaineering, (*b*) the experimental learning program on light, and (*c*) the criterion test. Each subject worked at his own speed. All subjects worked once through the learning program and were not allowed to turn back to earlier frames. Ten weeks after the learning and immediate test session the test was readministered to the two classes in order to provide a measure of long-term retention.

Results

Learning Response Difficulty. Scoring the responses to the learning program on light gives some indication of the difficulty level for the program responses. Subjects in the overt response group averaged 16% incorrect responses during learning. Subjects in the option group omitted an average of 12% of the responses and responded incorrectly to an average of 10% of the blanks. These error rates indicate that the program presented a relatively difficult level of responding as compared to typical continuous discourse programs.

Amount Learned. Measures of immediate and long-term retention were obtained from the scores on the test given immediately after the program on light and again 10 weeks later. A covariance analysis of the immediate test scores, using the reading scores from SCAT as the predictor variable, yielded

no significant differences among response modes ($F = 2.19$, $p < .15$, $df = 3/53$) or between classes. The average within-cell correlation between the reading scores and test scores was .54.

A corresponding analysis of covariance for the 10-week retention test data yielded significant response mode differences ($F = 3.39$, $p < .05$). Although there was an appreciable decrease in test scores for all groups, the relative standing of the four response mode means remained the same after the 10-week period, as shown in Table 6. The only notable change in the magnitude of the differences was the slightly larger superiority of the reading response group on the 10-week retention test.

Learning time. The mean learning times for the four response mode groups, shown in Table 7, differed significantly ($F = 26.08$, $p < .01$). These means show a decreasing trend corresponding to decreased explicit requirement to respond actively. Using the analysis of variance error term with $df = 54$ for t tests between means, it was found that both writing conditions (overt and option response modes) had significantly higher learning time means ($p < .01$) than both "thinking" conditions (covert and reading response modes). It is apparent that the outstanding factor in producing a learning time difference was the requirement to produce an overt response. There was also a significant learning time difference between classes ($F = 6.58$, $p < .05$). The class with the larger mean learning time had the higher mean test score.

TABLE 6

IMMEDIATE AND 10-WEEK RETENTION TEST SCORES ADJUSTED BY SCAT READING TEST: EXPERIMENT II

Response mode	Mean test score	
	Immediate	10-week retention
Overt	28.2	22.3
Option	27.3	21.8
Covert	25.3	20.3
Reading	29.5	25.3
Estimated *SD* within cells	4.6	4.3

TABLE 7

MEAN LEARNING TIMES (IN MINUTES) AND SCAT ADJUSTED EFFICIENCY SCORES: EXPERIMENT II

Response mode	Mean learning time	Mean efficiency (adjusted)
Overt	20.7	1.40
Option	19.8	1.38
Covert	13.9	1.95
Reading	11.5	2.81
Estimated *SD* within cells	3.5	0.71

Learning efficiency. The ratio of amount learned per unit of learning time was taken as an index of learning efficiency, as in Experiment I. A covariance analysis of efficiency scores produced a highly significant value of F for the response mode variable ($F = 13.39$, $p < .01$). The average within-group correlation between reading scores and efficiency scores was .45. Adjusted mean scores for response mode groups are shown in Table 7. The relative efficiency of response modes closely paralleled Experiment I.

Test time. It is appropriate to consider whether the experimental conditions had any effect on the time taken by subjects to complete the test. For the immediate test, there was a significant difference among response modes ($F = 4.18$, $p < .05$). The mean test times for the overt, option, covert, and reading response modes were 10.3, 12.0, 13.4, and 14.3 minutes, respectively. The order of these test times is exactly inverse to the order of learning times.

On the 10-week retention test, however, there were negligible time differences among response modes. The conditions of administration differed for the immediate and the 10-week retention test in that each subject began the immediate test as soon as he finished the learning program, while all subjects began the 10-week retention test at the same time. Since a reliable test time difference among response modes was found only for the immediate test, it is likely that the difference and the same finding in Experiment I is attributable to the time pressures associated with test administration. Subjects may have become aware that they were ahead of or behind the other members of the class as a result of learning time differences associated with response mode conditions, and some subjects may have compensated for this time difference by time spent on the immediate test. Apparently the group differences in time spent on the test had no measurable effect on immediate test score, inasmuch as an equivalent ordering of scores was obtained on the 10-week retention test for which there were no appreciable time differences.

DISCUSSION

The most general finding of Experiments I and II was the relatively low efficiency of the overt response condition and the relatively high efficiency of the reading condition for both factual and continuous discourse learning. The question of why the reading mode has been found to be more efficient can be answered in general by considering that the set for an effective pattern of learning can be established by presenting learning material in short logical steps (frames) that are interestingly written with an orientation for student responding. The further question of whether this mode of autoinstructional programming can maintain its high level of efficiency over long periods of time needs further investigation. Techniques used in the present study, such as underlining important words and promising a test at the end of a learning session may be sufficient to maintain learning efficiency. Other techniques, such as posing questions that will be answered by a block of subsequent frames and varying the format of the program, may be useful in maintaining efficiency.

In nearly every comparison on the amount learned measure, the performance of the overt response groups failed to exceed that of the reading groups. When learning responses were at the easiest level for the factual material, the inferiority of the overt response condition was most marked. The only evidence for a greater amount learned under the overt condition was found with intermediate response difficulty in Experiment I. The exceptionally long learning time taken by this high performing overt response group is consistent with the hypothesis presented earlier concerning the active participation that overt responding may stimulate at an intermediate level of response difficulty. Since about the same number of words were written at all difficulty levels, the additional learning time required by the overt group at the intermediate difficulty level cannot be attributed to time spent writing. The active participation evoked in this group might be described as searching or thought formulation.

The results obtained for independent frames showed that the value of a correct learning response is not so great as might be expected. The higher percentage of correct test responses associated with correct learning responses that is shown in Table 2 of the results can be attributed in part to the likelihood that some of the facts to be learned

were already known by some of the students. This likelihood is evidenced by the finding that the percentage of correct test responses for correct learning responses increased markedly as the amount of cueing was reduced. There also are differences in the inherent difficulty of the facts, increasing the likelihood that correct learning responses will be associated with correct test responses.

The most relevant item among the data for incorrect learning responses is the relatively high percentage of correct test responses for incorrect learning responses at the intermediate difficulty level. This is interpreted as indicating that cueing a response somewhat below threshold may be an effective learning condition if the errors thus elicited are followed immediately by corrective feedback. The failure of the option group to exceed the overt group in either amount learned or speed of completion of the task (Experiment II) also suggests that avoidance of errors during learning is not of critical importance.

The finding that the overt groups took much more time to learn nearly as much as the reading groups, might suggest that the overt groups overlearned the material and could be expected to retain it longer than the reading groups. The results on the 10-week retention test in Experiment II do not support this expectation. The superiority of the reading group was, in fact, slightly greater after the 10-week period.

It should be noted that the use of the test score and learning time ratio as an index of learning efficiency is not meant to imply a linear relationship between learning performance and time. Since, in fact, rate of learning tends to be higher early in learning, we might expect learning conditions with short learning times to give relatively high indices of efficiency. Unless one condition is superior to another in both amount learned and learning time, caution should be employed in generalizing relative efficiency to different learning times. In the studies reported, we do not have adequate data for evaluating the increased level of learning which might be achieved during the additional time made available by more "efficient" learning conditions.

A further consideration regarding learning efficiency is that a more precise measure of efficiency is obtained if amount learned is measured from a base determined by a chance test score or test scores obtained from a "no-learning" group. Such a refinement in the present studies would tend to magnify slightly the obtained superiority of the reading group in learning efficiency. Finally, it should be recognized, in general, that time is not the only basis for determining efficiency. Such items as amount or cost of learning material, energy expended, and motivation loss are possible alternative considerations that may assume considerable importance in practice.

REFERENCES

ANGELL, D., & LUMSDAINE, A. A. Stimulus presentation factors relevant to automated teaching. Paper read at Psychonomic Society, Chicago, September 1960.

BRIGGS, L. J. *A survey of cueing methods in education and in automated programs.* Pittsburgh: American Institute for Research, 1960.

COOK, J. O. Supplementary report: Processes underlying learning a single paired-associate item. *J. exp. Psychol.*, 1958, **56**, 455.

COOK, J. O., & KENDLER, T. S. A theoretical model to explain some paired-associate learning data. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research.* Washington, D. C.: National Academy of Sciences-National Research Council, 1956.

COOK, J. O., & SPITZER, M. E. Supplementary report: Prompting versus confirmation in paired-associate learning. *J. exp. Psychol.*, 1960, **59**, 275-276.

EVANS, J. L., GLASER, R., & HOMME, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book.* Washington, D. C.: National Education Association, 1960. Pp. 446-451.

GOLDBECK, R. A., & BRIGGS, L. J. An analysis of response mode and feedback factors in automated instruction. (Technical Report No. 2) Santa Barbara: American Institute for Research, 1960.

KLAUS, D. J., HOLMOK, E. S., LUMSDAINE, A. A., MEYER, D. S., & QUISENBERRY, K. S. *High school physics.* Vol. 2. *Reflection and refraction.* Pittsburgh: American Institute for Research, 1960.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

PRESSEY, S. L. Some perspectives and major problems regarding "teaching machines." In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book.* Washington, D. C.: National Education Association, 1960. Pp. 497-505.

SILVERMAN, R. E., & ALTER, M. Note on the response in teaching machine programs. *Psychol. Rep.*, 1960, **7**, 496.

SKINNER, B. F. Teaching machines. *Science*, 1958, **128**, 969-977.

(Received June 12, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 3, 119–126

CONCEPT ATTAINMENT BY DEAF AND HEARING ADOLESCENTS[1]

SOLIS L. KATES, LEE YUDIN, AND RONALD K. TIFFANY
University of Massachusetts
and Clarke School for the Deaf, Northampton, Massachusetts

The purpose was to investigate concept attainment by deaf and hearing adolescents. There were 30 deaf Ss matched first with 30 hearing Ss on sex, age, and IQ and then with 30 hearing Ss on sex, school achievement, and IQ. All Ss were tested individually on 6 concept problems. The conclusions were: (a) deaf Ss manifest capability equal to hearing Ss in tying together sensory impressions and in making use of these integrated sensory impressions to classify new objects, (b) deaf Ss show greater cautiousness in taking the initial step in problem solving, (c) deaf and hearing Ss proceed in concept attainment tasks by the use of similar strategies.

The purpose of this study is to investigate concept attainment by deaf and hearing adolescents in problem solving situations. Deafness renders difficult the coordination of speech categories to the nonlinguistic categories of the world (Brown, 1956). But where nonlinguistic categories have defining attributes based more upon sensory properties than upon socially prescribed properties, it would appear that the deaf child would not be inferior to the hearing child. If this latter notion is confirmed, then there would be reason to believe that deafness may not radically alter the nature of conceptual ability in deaf children, but would simply prevent the acquisition of categories, or make less available those categories based upon socialization.

Some studies indicate that the deaf subjects are less capable than hearing subjects of abstracting the essential attributes of objects or events that belong to a class and of using such attributes for identifying the new instances of this class without further learning (Hughes, 1959; McAndrew, 1948; Myklebust & Brutten, 1953; Oleron, 1953). A different view of the conceptual capacities of the deaf is stated by Kates, Kates, Michael, and Walsh (1961) and by Rosenstein (1959). For example, Kates et al. (1961) demonstrated that deaf subjects performed as well as hearing subjects, matched on age and IQ, when their categorizations were judged independently of their verbalizations.

The problem of this study was to determine how deafness affects first, the ability to attain conjunctive concepts (where all instances share common attributes) and second, the processes manifested in concept attainment (Bruner, 1957; Bruner, Goodnow, & Austin, 1956; Werner, 1948).

Specifically, the hypotheses of this study were:

1. There will be no differences in concept attainment between deaf adolescent subjects and older hearing control subjects matched on age and IQ in (*a*) the total number of choices before solution, (*b*) the latency of the first choice in each conceptual problem, (*c*) the total time taken to achieve solution, (*d*) the number of guesses about the correct concept, (*e*) the number and kind of changes made in attribute values

[1] This study was supported, in part, by a research grant (M-4392) from the National Institutes of Health, United States Public Health Service.

on the first choice, (*f*) the percentage of focusing changes made in the first four card choices, and (*g*) the comparison of their respective performances on the orderly and random array of stimulus materials.

2. Deaf adolescent subjects will be significantly superior in concept attainment to younger hearing control subjects matched on school achievement level and IQ in (*a*) the total number of choices before solution. Since the younger hearing subjects have also developed formal thinking processes (Inhelder & Piaget, 1958), it is hypothesized that there will be no differences in the underlying processes of concept attainment as follows: (*b*) the latency of the first choice in each conceptual problem, (*c*) the total time taken to achieve solution, (*d*) the number of guesses about the correct concept, (*e*) the number and kind of changes made in attribute values on the first choice, (*f*) the percentage of focusing changes made in the first four card choices, and (*g*) the comparison of their respective performances on the orderly and random array of stimulus materials.

METHOD

Subjects

One group consisted of 30 deaf students (17 girls and 13 boys) at the Clarke School for the Deaf where the oral method of training is employed. Each subject, profoundly deaf from birth or infancy, has no other physical defect, and is free from atypical personality characteristics. The deaf subjects have a mean IQ of 96.87 on the appropriate Wechsler intelligence scale, a mean age of 16.4 years, and a mean school grade achievement level on the Stanford Achievement Test equivalent to the 8.51 grade level.

Our first control group (C_1) consisted of 30 hearing subjects (17 girls and 13 boys) from the senior high school. They have a mean IQ (Wechsler and Otis) of 97.63, a mean age at 16.55 years, and were progressing at an average pace in their studies. Our second control group (C_2) was composed of 30 hearings subjects (17 girls and 13 boys) from the junior high school. Their average IQ (Wechsler and Otis) is 97.50, their average school achievement grade level (Stanford Achievement Test) is equivalent to the 8.50 grade level, and their average age is 13.6 years. All hearing subjects were free from overt personality and physical maladjustments.

There were no significant differences between the deaf subjects and the junior high school students on school achievement grade level and on IQ, the two characteristics on which these subjects were matched. No significant differences were found between deaf subjects and the senior high school students on the two relevant characteristics of age and IQ.

Concept Attainment Materials

The concept attainment materials consisted of 81 cards measuring 3″ × 5″, described by Bruner et al. (1956). On each card, there is a combination of four attributes comprising color, figures, number of figures, and borders. The color attribute has three values: red, black, and green. The figure attribute includes cross, square, and circle. The number attribute includes one, two, and three. There are one, two, or three borders, the three different values for the fourth attribute. Some of the cards are one black square with one border, three black squares with two borders, two green crosses with one border, three red circles with three borders.

The 81 cards were displayed in two different conditions. The first condition was an orderly one where all the cards were grouped systematically with similar attributes and values placed in juxtaposition (orderly board). The second condition was a random one with the 81 cards assigned to positions by the use of random numbers (random board).

Procedure

All subjects were tested individually on the same problems which were administered in the same order. The deaf subjects were divided into two subgroups of 15 subjects each, matched on IQ, school achievement grade level, and age. The older hearing control group was also divided into two subgroups of 15 subjects each, matched on age and IQ. The younger hearing control group was similarly separated into two subgroups

matched on school achievement grade level and IQ. One subgroup of deaf subjects, older hearing subjects, and younger hearing subjects, respectively, solved the first three conceptual problems on the orderly board (first half) and three problems on the orderly board again (the second half). The other subgroup of deaf subjects, older hearing subjects, and younger hearing subjects, respectively, solved the first three conceptual problems on the orderly board (first half) and three problems on the random board (second half). The nature of the task was completely described to the subjects who were to attain concepts that the experimenter had in mind; the concept was considered attained when the attribute values defining the concept were specified by the subjects.

All the subjects completed three practice problems before attempting the six experimental problems. The experimental problems required the subjects to attain the following concepts: in the first half, green crosses, two squares, circles and one border; and in the second half, red figures, three black figures, two squares and two borders. The dependent variables were:

1. The number of choices before attaining the concept.
2. The latency of the first choice.
3. The total time taken for attaining the concept.
4. The number of guesses made about the concept.
5. The number and kind of changes made in attribute values on the first choice. This is a measure of strategy and evaluates first how many attribute values the subjects change on their first choice as contrasted with the focus card and second whether the more abstract attribute (number) or the less abstract attribute (figure) is changed on the first card.
6. The percentage of focusing changes made in the first four choices. A focusing change is a change from the focus card in one value *only* of an untested attribute, thus evaluating the strategy pursued on the first four choices.

Statistical Procedure

Because we were not concerned with the overall differences among the three groups of subjects and because we were concerned with the differences between the deaf group and each matched hearing group separately, comparisons were made of the deaf subjects with each one of the two hearing control groups. The analysis of variance technique was used for all dependent variables.

RESULTS

Deaf and Older Hearing Subjects on Choices, Latency, Total Time, Guesses, Attributes Changed, and Focusing Changes. The results indicated there were no significant differences between the deaf subjects and the older hearing subjects on the number of choices necessary to solve the six conceptual problems (Table 1). The older hearing group showed a significantly (at .025 level) smaller latency for their first choices than the deaf subjects (Table 1). In the total time taken to solve the conceptual problems, there was no significant difference between the deaf and older hearing groups. The older hearing subjects offered a significantly (at .025 level) greater number of incorrect guesses about the defining attributes before solution than did the deaf subjects (Table 1).

There were no significant differences between the two groups on the number of changes and on the kinds of changes in attributes on the first card choice and on the percentage of focusing changes made in the first four chioces (Table 2).

Deaf and Older Hearing Groups on the Orderly-Orderly and Orderly-Random Sequences. There were no significant differences between deaf and older hearing subjects with regard to the variables of card choices, latency, total time, guesses, number and kinds of attributes changed on first choice, and percentage of focusing changes on first four choices when their respective performances were compared on the orderly-orderly and the orderly-random sequences (Tables 1 and 2). Finally, the differences between the respective

TABLE 1

MEANS, *SD*s AND *F* TESTS FOR DEAF AND TWO CONTROL GROUPS ON CHOICES, LATENCY, TIME, AND GUESSES[a]

Group	Choices		Latency (in seconds)		Total Time (in seconds/10)		Guesses	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
All problems								
Both sequences (O-O O-R)								
C_1 (older hearing group)	61.4	25.2	39.4[b]	9.8	118.3	65.8	22.6[b]	9.3
D (deaf group)	61.7	20.8	49.3	20.6	128.5	62.8	16.3	9.2
C_2 (younger hearing group)	73.1[c]	26.9	35.2[d]	17.3	103.2	71.3	19.3	8.6
Orderly-orderly sequence (O-O)								
C_1	61.0	17.0	35.2	7.4	117.0	68.9	24.4	11.0
D	59.6	19.9	50.6	19.4	131.3	58.2	14.3	4.9
C_2	68.4	19.5	35.3	15.6	94.6	32.7	20.6	7.9
Orderly-random sequence (O-R)								
C_1	61.8	31.4	43.6	10.1	119.6	62.4	20.9	6.9
D	63.8	21.5	48.0	21.6	125.6	67.0	18.4	6.1
C_2	77.8	32.0	35.0	18.9	111.8	94.6	18.0	9.0
First half								
Orderly first half of O-O sequence								
C_1	33.2	11.6	18.8	4.7	70.0	47.3	14.6	6.5
D	31.8	15.8	26.6	10.7	75.3	47.6	7.3	3.7
C_2	30.4	13.2	19.6	12.1	42.8	22.2	9.4	4.0
Orderly first half of O-R sequence								
C_1	32.2	18.4	22.5	7.5	70.1	46.8	11.6	5.2
D	33.3	14.9	20.6	9.2	69.5	54.2	10.2	9.0
C_2	37.4	19.1	18.3	8.3	54.9	43.8	8.8	5.0
Second half								
Orderly second half of O-O sequence								
C_1	27.8	9.6	16.4	4.7	47.0	28.1	10.1	4.7
D	27.8	9.3	24.0	10.2	56.0	42.2	7.0	3.0
C_2	38.0	15.1	15.6	7.2	51.7	32.3	11.2	6.1
Random second half of O-R sequence								
C_1	29.6	19.6	20.3	10.2	49.5	31.1	9.2	4.1
D	30.5	12.6	27.3	17.3	56.1	27.1	8.1	3.2
C_2	40.4	18.9	16.7	11.6	56.9	58.9	9.1	4.7

[a] All the *F* tests for the comparisons between the corresponding means of the deaf group and the older hearing group and between the corresponding means of the deaf group and the younger hearing group were not significant unless otherwise indicated.

[b] *F* test was significant for the comparison between the means of the deaf group and older hearing group.

[c] *F* test showed a trend at .10 level for the comparison between corresponding means of the deaf group and the younger hearing group.

[d] *F* test was significant for the comparison between the means of the deaf group and the younger hearing group.

TABLE 2

MEANS, *SD*s, AND *F* TESTS FOR DEAF AND TWO CONTROL GROUPS ON FIRST CHOICE CHANGES AND FOCUSING CHANGES[a]

Group	Number of attribute changes on first choice		Number of number changes on first choice		Number of figure changes on first choice		Cumulative percentage of focusing changes in first four choices	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
All problems								
Both sequences (O-O O-R)								
C_1 (older hearing group)	7.7	1.6	1.7	1.3	1.6	1.2	3.7	.6
D (deaf group)	7.5	1.5	2.2	1.2	1.1	1.2	3.5	.6
C_2 (younger hearing group)	7.4	1.4	2.2	1.3	1.2	1.1	3.7	.5
Orderly-orderly sequence (O-O)								
C_1	7.1	.4	1.8	1.3	1.4	1.3	3.9	.5
D	7.2	1.1	2.2	1.2	1.2	.6	3.6	.4
C_2	7.2	1.4	2.6	1.2	1.1	.8	3.8	.5
Orderly-random sequence (O-R)								
C_1	8.3	1.8	1.6	1.3	1.7	1.1	3.4	.7
D	8.1	1.7	2.1	1.2	1.0	1.3	3.4	.7
C_2	7.6	1.3	1.8	1.2	1.3	.7	3.6	.7
First half								
Orderly first half of O-O sequence								
C_1	3.4	.6	.9	.9	.4	.6	1.9	.3
D	3.5	.8	1.2	.7	1.2	.7	1.9	.3
C_2	3.5	.6	1.3	.9	1.3	.8	1.9	.3
Orderly first half of O-R sequence								
C_1	3.6	1.1	.6	.7	.7	.7	1.8	.2
D	3.5	.8	.8	.7	.8	.7	1.8	.1
C_2	3.3	.4	1.0	.8	1.0	.8	1.9	.2
Second half								
Orderly second half of O-O sequence								
C_1	3.1	.9	.9	.9	1.0	.9	2.0	.3
D	3.7	1.5	1.0	.9	.7	.8	1.7	.3
C_2	3.1	1.0	1.2	.9	.7	.7	1.9	.2
Random second half of O-R sequence								
C_1	4.6	1.4	.9	.7	1.0	.2	1.5	.4
D	4.3	1.4	1.3	.7	.5	1.0	1.5	.6
C_2	4.2	1.0	.8	.6	1.0	.9	1.7	.3

[a] All the *F* tests for the comparisons between the corresponding means of the deaf group and the older hearing group and between the corresponding means of the deaf group and the younger hearing group were not significant.

performances of the deaf and the older hearing subjects on the first half and on the second half of the orderly-orderly sequence, and on the first half and the second half of the orderly-random sequence were not significantly different for card choices, latency, total time, guesses, number and kind of attributes changed on the first choice, and percentage of focusing changes on the first four choices (Tables 1 and 2).

Deaf and Younger Hearing Subjects on Choices, Latency, Total Time, Guesses, Attributes Changed, and Focusing Changes. The results indicated that the deaf subjects and the younger hearing subjects did not differ significantly on the number of choices required before solving the six conceptual problems. However, there was a trend since the deaf subjects required a smaller number of choices than the younger hearing group with this difference reaching the .10 level of significance (Table 1).

The younger hearing subjects took significantly (at .01 level) less time to make their first choices in the conceptual problems than the deaf subjects (Table 1). The deaf and younger hearing subjects were not significantly differentiated on total time, number of guesses, number and kind of attributes changed on first choice, and on the percentage of focusing changes on the first four choices (Tables 1 and 2).

Deaf and Younger Hearing Groups on the Orderly-Orderly and Orderly-Random Sequences. There were no significant differences between deaf and younger hearing subjects with regard to card choices, latency, total time, guesses, number and kind of attributes changed on first choice, and percentage of focusing changes on the first four choices when their respective performance were compared on the two different sequences, the orderly-orderly sequence and the orderly-random sequence. Finally, there was no significant finding that the deaf and the younger hearing subjects were different when their respective performances on the first half of each sequence were contrasted with their performances on the second half of each sequence for the variables of card choices, latency, total time, guesses, number and kind of attributes changed on first choice, and percentage of focusing changes on the first four choices (Tables 1 and 2).

DISCUSSION

The results confirm Part *a* of the first hypothesis; the deaf and the older hearing subjects are equally capable of making use of discriminable attributes to distinguish exemplars of a concept from nonexemplars. Here, the deaf manifest capability equal to the hearing, in integrating the sensory impressions that comprise the desired concept and in making use of these integrated sensory impressions to classify new objects. It can also be inferred that language achievement (the older hearing subjects are assumed to be more proficient in language than the deaf subjects) does not have important significance for efficiency in attaining concepts based upon sensory similarity rather than upon similarity dictated by culture or convention.

That part of the first hypothesis involving the latency of the first choice in each conceptual problem was not supported. Rather than no differences being revealed, it appeared that the deaf took significantly more time in making this first choice than the hearing. It may be that this greater latency in responding to problems may be a function of the sensory deprivation that the deaf suffer, which finds some expression in the greater reluctance to engage in initial exploratory behavior. But that

part of the first hypothesis dealing with time was substantiated, since both hearing and deaf subjects required about the same time to solve the problems.

The older hearing group offered a significantly greater number of guesses than the deaf thus invalidating that part of the first hypothesis dealing with guesses. Explanations for this significant difference may be the recognition by the deaf of their relative inability in speech communication and their consequent reluctance to offer spoken guesses, and their possibly greater cautiousness and fear of failure prior to embarking upon a course of action.

The parts of the first hypothesis dealing with the number and kind of changes made in attribute values on the first choice and with the percentage of focusing changes made in the first four card choices were confirmed. No significant differences in these two measures yield support to the notion that deaf and older hearing groups pursue similar strategies in solving conceptual problems.

The type of changes on the first choice—whether the figure or the number is changed—would seem to be a measure of ability to deal with either a relatively less abstract attribute, the figure, and/or a relatively more abstract attribute, the number. When the subject changed the number, this signified that he was able to detach the number from the focus card, an attribute more difficult to abstract (Heidbreder, 1946). On the other hand, the change of the figure signified that a more concrete attribute was abstracted for change. The deaf and the older hearing subjects showed no differences in the kinds of changes made on the first card. It would appear that the deaf are no more attracted to the more concrete attributes than the hearing, and are able to manipulate the more abstract attributes.

Finally, the presumably greater difficulty in concept attainment when the cards are in random, rather than in orderly, array (found by Bruner et al., 1956) did not distinguish the deaf and the older hearing groups on any variable. In the Bruner et al. (1956) procedure, one group of subjects was introduced immediately to the random board, and there was a decrement in their performance as compared to the performance of a matched group on the orderly board. However, in our experiment, there were no differences apparently because performance on the orderly board preceded the performances on both orderly and random boards and resulted in modes of attaining concepts that carried over with equal proficiency to these boards.

Except for two parts, the second hypothesis was confirmed, indicating that the deaf and younger hearing groups were not different on total time, number and kind of changes made in attribute values on the first choice, percentage of focusing changes made in the first four card choices, and comparison of their respective performances on the orderly and randomly arrayed stimulus materials. Hence, with regard to strategy and to the effect of increasing difficulty in stimulus materials, both deaf and younger hearing groups manifest about similar performance.

The part of the second hypothesis dealing with the superiority of the deaf in number of choices required for solution over the younger hearing was not substantiated. Nevertheless, the trend was in the predicted direction (at the .10 level) with the deaf requiring a smaller number of choices than the younger hearing group.

The deaf group have a longer latency on their first choices than the younger hearing group, and thus that part of the second hypothesis was not supported.

It should be recalled that the deaf have significantly longer latencies than the older hearing group. These two findings point to the condition of deafness or to the training given to the deaf group as possibly inhibiting initial exploratory behavior (Bruner, 1961).

No significant differences between the deaf and the younger hearing group in number and kind of changes on first choice and in percentage of focusing changes on first four choices point to a similarity in their concept attainment strategies as predicted in our second hypothesis. At approximately 14 years of age, adolescents appear to develop formal operations that permit them to deal with problems using the same processes as older adolescents of the same IQ (Inhelder & Piaget, 1958). The measures of number and kind of attribute changes on first choice and of percentage of focusing changes on first four choices showed significant variations (at .05 level) from the first half to the second half of the experiment in the direction predicted by increased difficulty of task. Hence, failure to find differences between the deaf and the two hearing groups, respectively, is probably not due to the lack of sensitivity of these measures.

REFERENCES

BROWN, R. Language and categories. In J. S. Bruner, J. J. Goodnow, & G. A. Austin, *A study of thinking*. New York: Wiley, 1956. Pp. 247-312.

BRUNER, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123-152.

BRUNER, J. S. The cognitive consequences of early sensory deprivation. In P. Solomon, P. Kubzansky, & P. H. Leiderman (Eds.), *Sensory deprivation*. Cambridge, Mass.: Harvard Univer. Press, 1961.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.

HEIDBREDER, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191-223.

HUGHES, R. B. A comparison of verbal conceptualization in deaf and hearing children. Unpublished doctoral dissertation, University of Illinois, 1959.

INHELDER, BARBEL, & PIAGET, J. *The growth of logical thinking from childhood to adolescence*. New York: Basic Books, 1958.

KATES, S. L., KATES, W. W., MICHAEL, J., & WALSH, T. M. Categorization and related verbalizations in deaf and hearing adolescents. *J. educ. Psychol.*, 1961, **52**, 188-194.

MCANDREW, H. Rigidity and isolation: A study of the deaf and the blind. *J. abnorm. soc. Psychol.*, 1948, **43**, 476-494.

MYKLEBUST, H., & BRUTTEN, M. A study of the visual perception of deaf children. *Acta Oto-laryngol. Suppl., Stockh.*, 1953, No. 105.

OLERON, P. Conceptual thinking of the deaf. *Amer. Ann. Deaf*, 1953, **98**, 304-310.

ROSENSTEIN, J. Perception and cognition in deaf children. Unpublished doctoral dissertation, Washington University, 1959.

WERNER, H. *Comparative psychology of mental development*. (Rev. ed.) New York: International Univer. Press, 1948.

(Received November 14, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 3, 127–131

A COMPARISON OF TWO RESPONSE MODES IN AN AUTOINSTRUCTIONAL PROGRAM WITH CHILDREN IN THE PRIMARY GRADES

EVAN R. KEISLAR AND JOHN D. McNEIL
University of California, Los Angeles

To test the value of overt responding with young children, a total of 198 primary grade children divided into 2 matched groups were individually taught, by a teaching machine, a 3-week unit in physical science. The overt response group was required to respond to each frame by selecting the correct answer before the program would advance. The nonovert response group merely observed the entire program as it was automatically presented. Both groups learned significantly more than an uninstructed control group, but the difference between nonovert and overt responders was not significant.

A number of recent studies have failed to find any advantage to the procedure of having learners respond overtly to the frames in an autoinstructional program. Permitting learners to "think" or simply to read the frame with the correct answer inserted produced results which were usually effective and sometimes even superior (Evans, Glaser, & Homme, 1960; Goldbeck, 1960; Roe, Massey, Weltman, & Leeds, 1960; Silberman, Melaragno, & Coulson, 1961; Silberman, Melaragno, Coulson, & Estevan, 1961; Silverman & Alter, 1961; Zuckerman, Marshall, & Groesberg, 1961). For example, Goldbeck found with junior high school students that subjects who read the frames, the Implicit Response Group, learned more, with easy items at least, than those who were required to write these responses, the Overt Response Group. When the time required to learn was taken into account, this superiority was even greater. These studies, however, used fairly mature students as subjects, individuals who had probably already learned how to learn.

Using young children as subjects with a program which called for a choice between two responses on each frame, Suppes and Ginsberg (1960) obtained different results. They found that 5- and 6-year-old children who were required to correct overtly their errors learned more than children who were merely informed whether they had responded correctly or not.

It should be noted that most investigations with both older and younger subjects measured effects from relatively short periods of instruction, usually a single session. However, Holland (1960), using a semester-long program with college students found that overt responding to incomplete statements was more effective than simply reading the completed statements. It is important to know the effect of length of program and the conditions, if any, which make overt responding of young children desirable.

It was the purpose of this study to test the value of overt responding with young children using a program which required approximately 3 weeks of instruction. Specifically, it was hypothesized that children who are required to respond to each frame of a multiple-choice program with the correct answer will learn more than children who are not required to respond but who

observe the program and are also given at each step information as to the correct answer.

It was assumed that both those children who were required to respond overtly as well as those who were not so required would engage in implicit behavior. The purpose of the study was not to assess the value of implicit behavior; and hence, no attempt was made to inhibit covert responding of any child. The criterion posttest was composed largely of new questions which could only be answered if the subject had acquired appropriate implicit responses. A large portion of the multiple-choice frames in the program, therefore, also required this mediating behavior. The overt response of pushing buttons was not the terminal behavior sought. However, the very fact that the Overt Response Group was required, in connection with each frame, to make a visible overt act—to be scored right or wrong—was judged to increase the probability of the subject's producing the desired implicit behavior. Speaking popularly, the subjects with an investment openly at stake should pay closer attention and hence, learn more than those who make no visible commitment.

METHOD

Subjects

The subjects for this investigation consisted of 118 children from School A in Grades 1 and 2, and 80 children in School B from Grades 1 and 3. At each school the children were divided by matched pairs into two groups, an Overt Response and a Nonovert Response Group. In School B, 57 additional uninstructed children were used as a control group, permitting an estimate of the effectiveness of the program.

Procedure

All of the subjects from both experimeental groups were taught individually by a device, the videosonic teaching machine (experimental model), which projected slides and played a commentary to the learner through earphones. On the response panel, three buttons were available for responding to questions, each of which required a selection from two or three alternatives. The program consisted of 432 multiple-choice frames dealing with physical science and requiring each child to spend approximately 20 minutes a day for 12 days. The difficulty level of this program, which has been described elsewhere (Keislar & McNeil, 1961) was previously found to be approximately 14% (proportion of frames missed on first attempt) for a comparable sample of first-grade children.

Children in the Overt Response Group were required to press the correct button, the one in front of the appropriate picture, before the program would advance. When they pressed the correct button, a green light was always turned on in front of the right answer. If they selected the wrong picture, these children were required to press a special button to clear the keyboard so as to permit another attempt.

Children in the Nonovert Response Group were individually instructed by the same device, but these children were not required to press buttons or make any overt response. To these "passive" observers the same program was automatically presented. A 5-second pause followed each question, approximating the time required by the other group to answer. At the end of this 5-second interval, a green light was turned on automatically in front of the right answer.

In School A, 12 teaching machines were used, six for the Overt Response Group in one room, and six for the Nonovert Response Group in another room. In School B, 10 machines were used in different parts of the same room, five for each of the two groups. Screens were placed between the machines to minimize for each subject distractions from other children.

Although the children in the Nonovert Response Group were given no instructions to respond, on occasion a few children spontaneously pointed to a picture to indicate their answers to the questions they heard from the tape, but the amount of such overt responding decreased during the 3-week period.

Posttest

In School A, the posttest consisted of, first, a standardized interview in which children were individually asked to explain or describe

certain physical phenomena, and, second, a multiple-choice group test which required each child to make a mark on his paper in a square corresponding to the correct picture projected on a screen. In School B, the questions from the standardized interview were placed in multiple-choice form using as alternatives responses given earlier by children in the other school. In addition, the same multiple-choice test given in School A was used. In School B, therefore, both parts of the posttest consisted of multiple-choice items.

In both schools each part of the posttest called for the application of principles taught. For example, children were asked to explain the formation of water on the outside of a glass of ice water, a task requiring a knowledge of molecular movement and condensation. The subjects were given the first part of the posttest on the school day following the completion of the program; the second part of the posttest was administered on the next school day.

RESULTS

The reliability of the posttest in each school was computed by a split-half method, yielding, after correction, indices of .86 and .87 for the two schools based on all subjects in both experimental groups. These relatively high figures are, in part, attributable to the large spread of scores on the posttest. Such wide individual differences in scores are to be expected from a test of transfer; a more restricted range would have been expected if the test items had been identical with those in the program (McNeil & Keislar, 1961).

Since the criterion test was slightly different in the two schools, the results are considered separately for each school, as shown in Table 1. It may be seen that the mean score for the Overt Response Group is only slightly higher in each school. In both instances, the difference between the mean scores of the Overt and Nonovert Response Groups is practically zero, and is far from being statistically significant.

In School B, a control group of 57 youngsters from Grades 1 and 3 appeared each day to study a mathematics program. This group showed an average score on the multiple-choice posttest of 36.6 with a standard deviation of 6.6. Since the differences between the experimental groups and the control group are both highly significant, it is clear that the science program was effective in teaching.

DISCUSSION

Although the results of this investigation are consistent with many studies involving older learners, they appear at variance with the findings of Suppes

TABLE 1

POSTTEST RESULTS IN TWO SCHOOLS FOR OVERT RESPONSE AND NONOVERT RESPONSE GROUPS

School	Group	N	M[a]	SD	M_1-M_2	$SE_{M_1-M_2}$[b]
A	Overt Response	59	39.3	10.5	.1	1.4
A	Implicit Response	59	39.2	11.1		
B	Overt Response	40	49.0	8.0	.9	1.2
B	Implicit Response	40	48.1	9.1		

[a] Scores for School A were obtained from a somewhat different test than the one used in School B.

[b] These standard errors are based on differences between pairs matched for grade, sex, intelligence, and, in addition, for School A only, pretest score.

and Ginsberg (1960). There is, however, an important difference between this latter study and the present one. In the Suppes and Ginsberg experiment, children who were not required to make an overt response of correction were merely told that their response was incorrect. These children did not do as well as those who were required to give the correct response. Where there are only two alternatives, it might be expected that subjects, when told their answer is wrong, will "infer" that the other alternative is correct. Although this may be true for adults, children with limited experience may not do this. Young subjects may need to be told that the other alternative is correct. In the present investigation, children in the Nonovert Response Group were always given information for each frame telling which response was correct; they were, therefore, given in this way the same "information" as the children in the Overt Response Group.

The failure to find a difference in favor of overt responding must be interpreted in the light of a number of considerations:

1. The responses made by the Overt Response Group were not those measured by the posttest. To evaluate the function of overt responses more effectively, it may be necessary to demonstrate that the behavior required on the posttest is directly related to the overt response demanded during the program. For example, if one were attempting to teach children to pronounce new words, it is conceivable that requiring the children to make an overt response, like saying words in completion of sentences, would be more effective than permitting the children merely to hear the completed sentences spoken.

2. The definition of an overt response adopted by this study is restricted to the required visible registration of the subjects' answers. Assessment of the influence of only this overt behavior was attempted.

3. A report of "no difference" needs to be viewed in terms of the precision of the measures. The reliability of the criterion test seems to be satisfactory. The highly significant difference of 13 points between the two experimental groups and the control group is evidence that a measurable product was effectively taught. The standard errors of the differences between the two experimental groups (1.4 and 1.2 for Schools A and B, respectively) suggest that one may reject at the .01 level the hypothesis that the true difference is greater than four points in either school. If a four-point difference is judged to be unimportant, the hypothesis of a "negligible" difference may be accepted.

4. The difficulty level of the program used may be too high to permit the Overt Response Group to show up well. Because this group was required to respond before being shown the right answer to each item, with difficult items they may have made more errors than those in the Nonovert Response Group. These latter children may well have waited until the green light showed the right answer before making an assuredly correct implicit response. Errors made by children during a program, even if corrected, may result in inferior learning.

5. For the learning of the new patterns of implicit responses called for in this study, the required overt response involving a reinforcement or "stake" may add nothing to the implicit behavior of the learner. The program itself, featuring color, sound, movement, and presentation of relevant cues, all factors involved in good programing, may be the critical factor. In short, the elicitation of appropriate implicit behavior may have been accomplished

equally well, at least for this 3-week period of instruction, in both the Overt and Nonovert Response Groups because of the stimulus variables in the program.

REFERENCES

EVANS, J. L., GLASER, R., & HOMME, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Education Association, 1960. Pp. 446-451.

GOLDBECK, R. A. The effect of response mode and learning material difficulty on automated instruction. Report No. 1, September, 1960, Contract No. Nonr-3077(00), American Institute for Research, Pittsburgh.

HOLLAND, J. G. Program design and use. Paper read at American Psychological Association, Chicago, September 1960.

KEISLAR, E. R., & MCNEIL, J. D. Teaching scientific theory to first grade pupils by auto-instructional device. *Harv. educ. Rev.*, 1961, **31**, 73-83.

MCNEIL, J. D., & KEISLAR, E. R. Individual differences and effectiveness of autoinstruction at the primary grade level. *Calif. J. educ. Res.*, 1961, **12**, 160-164.

ROE, A., MASSEY, M., WELTMAN, G., & LEEDS, D. Automated teaching methods using linear programs. Report No. 60-105, December 1960, Department of Engineering, University of California, Los Angeles.

SILBERMAN, H., MELARAGNO, R. J., & COULSON, J. Confirmation and prompting with continuous discourse material. *Psychol. Rep.*, 1961, **8**, 235-238.

SILBERMAN, H., MELARAGNO, R. J., COULSON, J., & ESTEVAN, D. Fixed sequence versus branching autoinstructional methods. *J. educ. Psychol.*, 1961, **52**, 166-172.

SILVERMAN, R. E., & ALTER, M. Response mode, pacing, and motivational effects in teaching machines. Technical Report No. 507-3, June 1961, United States Naval Training Device Center, Port Washington, New York.

SUPPES, P. & GINSBERG, R. Application of a stimulus sampling model to children's concept formation of binary numbers with and without an overt correction response. Technical Report No. 35, December 1960, Institute for Mathematical Studies in the Social Sciences, Stanford, California.

ZUCKERMAN, C. B., MARSHALL, G. R., & GROESBERG, S. Research in the automation of teaching. Technical Report No. 661-1, February 1961, United States Naval Training Device Center, Port Washington, New York.

(Received November 14, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 3, 132–143

THE PREDICTION OF THE ACADEMIC, ARTISTIC, SCIENTIFIC, AND SOCIAL ACHIEVEMENT OF UNDERGRADUATES OF SUPERIOR SCHOLASTIC APTITUDE[1]

JOHN L. HOLLAND AND ALEXANDER W. ASTIN
National Merit Scholarship Corporation

Undergraduate achievement in academic, artistic, scientific, and social areas was predicted from extensive assessments in the high school senior year for 4 independent samples of high aptitude students (N ranged from 354 to 953) over intervals from 1 to 4 years. The criteria of college achievement included grades, checklists of unusual accomplishments in science and art (e.g., publications, winning public recognition through awards and prizes), and election to student offices. Predictors included: scholastic aptitude, personality, originality, interest, parental attitude variables, and background information. Correlational analyses indicated that achievers in each of the 4 areas resemble our stereotypes of the scientist, artist, leader, and academic achiever.

Previous reports on the superior student have been concerned with the prediction of college freshmen grades (Holland, 1958, 1959, 1960), and with the correlates of scientific, artistic, and academic achievement at the high school level (Holland, 1961). The present study, an extension of these four earlier studies, relates the student's attributes as manifested in high school, to his achievement in college over 1-, 2-, 3-, and 4-year intervals.

Four different kinds of talented performance served as the criteria of undergraduate achievement. They were: (*a*) creative performance in the arts (artistic achievement), (*b*) creative performance in the sciences (scientific achievement), (*c*) high freshman grades in college (scholastic achievement), and (*d*) leadership (social achievement). Although scholastic achievement was reported on in earlier papers, it is discussed again here for the purpose of comparing it with the three other kinds of achievement.

The earlier studies presented extensive information about the student's background and attributes. The purpose of the present paper is to organize this information so as to reveal what attributes are associated with each different kind of achievement. It was assumed that the formulations derived from these correlational clusters would be useful in developing preliminary statements of a theory of achievement.

Most of the previous studies of achievement at the undergraduate level have been limited to scholastic achievement. However, it has been found that getting good grades in college has little connection with more remote and more socially relevant kinds of achievement; indeed, in some colleges, the higher the

[1] This study is a part of the research program of the National Merit Scholarship Corporation and was supported by grants from the National Science Foundation and the Old Dominion Foundation. We are indebted to T. R. McConnell for providing some of the predictor data for the 1956 sample, and to the following persons for their assistance throughout this project: Virginia Chalmers, Laura Kent, and Elizabeth van Laer.

student's grades, the less likely it is that he is a person with creative potential (Holland, 1961). So it seems desirable to extend our criteria of talented performance.

METHOD

Student Samples

The four student samples were obtained from the four studies reported earlier (Holland, 1958, 1959, 1960, 1961). Although these National Merit finalists were originally drawn at random from pools of the 5,000 to 10,000 students who represented by state the top segments in aptitude of participants in the National Merit program (who numbered 58,000 to 600,000 for the years 1956 and 1959), the samples used in the present study cannot be called representative since only those students for whom we had complete predictive and criterion data were included. These students can be best characterized as students of extremely high scholastic aptitude; the *N*s, means, and *SD*s for the four samples are shown in Table 1.

This summary suggests that the four samples are comparable, at least with regard to aptitude, and that they do not differ much from the original samples from which they were drawn (see earlier reports).

Predictors

The predictors consisted of variables selected from the original assessments and obtained when the students were high school seniors. For the 1956 sample, these variables included the Originality, Complexity-Simplicity, Psychopathic Deviate, and Hypomania scales of the Omnibus Personality Inventory (OPI), and the Aviator, Physicist, Social Science Teacher, Accountant, Sales Manager, Musician, Occupational Level, and Masculinity-Femininity scales of the Strong Vocational Interest Blank (SVIB).

For the 1957 sample, the predictors included all the scales of the California Psychological Inventory (CPI), and the Inventory of Beliefs (IB).

For the 1958 sample, the predictors included the Cyclothymia (A), Dominance (E), Surgency (F), Autia (M), Radicalism (Q_1), and Self-Sufficiency (Q_2) scales of the Sixteen Personality Factor Questionnaire (16 PF); the Physical Activity, Intellectuality, Responsibility, Conformity, Verbal Activity, Emotionality, Control, and Status scales of the Vocational Preference Inventory (VPI); and the Dominance, Dependency, Play, Persistence, Super-Ego and Tolerance for Ambiguity scales of the National Merit Student Survey (NMSS).

For the 1959 sample, the predictors included the Complexity-Simplicity, Independ-

TABLE 1

SCHOLASTIC APTITUDE (SAT) LEVELS OF STUDENT SAMPLES

Sample	SAT	Boys		Girls	
		X	*SD*	*X*	*SD*
1956[a]	Verbal	691.40	53.85	707.27	44.72
	Math	726.36	60.00	673.37	74.83
1957[b]	Verbal	673.35	42.76	687.14	38.12
	Math	714.38	68.04	660.14	77.64
1958[c]	Verbal	680.91	54.59	689.04	61.36
	Math	712.03	68.60	650.96	72.54
1959[d]	Verbal	658.63	57.15	656.70	59.23
	Math	700.03	60.73	638.25	66.44

[a] Boys, $N = 369$; girls, $N = 154$.
[b] Boys, $N = 681$; girls, $N = 272$.
[c] Boys, $N = 294$; girls, $N = 128$.
[d] Boys, $N = 282$; girls, $N = 172$.

ence of Judgment, and Originality scales of Barron's Inventory of Personal Philosophy; the Mastery and Deferred Gratification scales of the NMSS; the Seclusion of the Mother, Excluding Outside Influences, Avoidance of Communication, Ascendency of Mother, Intrusiveness, and Acceleration of Development scales of the Parental Attitude Research Instrument (PARI); the Initiative Scale of the Ghiselli Self-Description Inventory; father's rank-ordering of nine traits for his children; and the Intellectuality, Responsibility, Emotionality, Control, and Status scales of the VPI.

A number of other predictors were common to two or more of the four student samples: the Mathematics and Verbal scales of the Scholastic Aptitude Test (SAT), high school grades (HSR), self and teacher ratings, the Creative Activities Score (CAS), and scores for creative performance in science and art during the high school years.

In all, from 24 to 46 predictors were correlated with the four criteria of undergraduate achievement for each student sample. Complete descriptions of these predictors are contained in the earlier reports.

Criteria of Achievement

A record of each student's college performance was obtained by polling the four samples during the spring of 1960. At that time, the 1956 sample was at the end of its senior year in college (4-year interval between predictors and criteria), the 1957 sample was at the end of its junior year (3-year interval), the 1958 sample was at the end of its sophomore year (2-year interval), and the 1959 sample was at the end of its freshman year (1-year interval). Students filled out the form by checking those achievements which applied to them. The following items from the form comprise the achievement scales:

Leadership

1. Elected to one or more student offices. Circle appropriate number: 1 2 3 4 5 6 7 8 9 or more.
2. Initiated a business enterprise of any kind. Explain.
3. Initiated a student or social organization of any kind. Explain.
4. Received an award or special recognition for leadership of any kind. Explain.

Scientific Achievement

1. Received a research grant or National Science Foundation Fellowship. If grant, indicate nature.
2. Received a prize or award for scientific paper or project.
3. Gave an original paper at a scientific or professional meeting sponsored by a professional society or association. Name of society or association.
4. Had scientific or scholarly paper published (or in press) in a scientific or professional journal. Give complete reference here if published.
5. Invented a patentable device. If yes, indicate briefly its nature.
6. Constructed an original piece of scientific apparatus. Describe briefly.

Artistic Achievement

1. Won one or more speech or debate contests.
2. Had minor roles or leads in plays not produced by a college or university. Give name of organization producing play.
3. Was regular performer on radio or TV program.
4. Won prize or award in a musical competition as a performer.
5. Composed music which has been given at least one public performance.
6. Arranged music for public performance.
7. Had one or more musical publications.
8. Performed with a professional orchestra.
9. Gave public recital (not collegiate).
10. Received a literary award or prize for creative writing (poem, novel, play, etc.).

Academic Achievement (grades in freshman year) equals the Honor Point Ratio (HPR) where A = 4, B = 3, C = 2, D = 1, and F = 0. Table 2 shows the intercorrelations of these criteria for the four student samples. The criteria appear to be relatively independent, except for the Artistic and Social Achievements, which are consistently positively related in all samples.

RESULTS

The correlations between the student's attributes as manifested in high

TABLE 2

INTERCORRELATIONS OF THE FOUR CRITERIA OF STUDENT ACHIEVEMENT IN COLLEGE FOR FOUR GROUPS OF STUDENTS

Criterion	Academic				Artistic				Social				Scientific			
	1956	1957	1958	1959	1956	1957	1958	1959	1956	1957	1958	1959	1956	1957	1958	1959
Academic					−05	−06	−06			06	08	18		13	08	−02
Artistic	08	08	01						25	22	22	23	−04	04	−01	−01
Social	11	12	14		31	21	21	36					07	04	04	00
Scientific	−01	13	00		−15	04	08	08	−05	00	−14	03				

Note.—Correlations for boys are shown in upper right triangle of matrix for 1956, 1957, 1958, and 1959 samples. Corresponding correlations for girls are shown in lower left triangle.

school and his subsequent achievements in college are presented in four tables (Tables 3-6).[2] To make these extensive findings more comprehensible, personality summaries were prepared for each kind of achiever: the Social, Artistic, Scientific, and Scholastic Achiever. The variables which support each statement are given in the parentheses following the statement. Interpretative or speculative statements can be identified by their lack of variable numbers. Nearly all the relationships are predictive, i.e., from high school to college, although for some of the samples certain other personal variables such as self-ratings were collected at the same time as the criterion data; in the tables, concurrent relationships are indicated by a "c" after the correlation.

For the most part, each descriptive summary is limited to findings which are either replicated (i.e., significant correlations are found between *identical* predictors and the criterion) in two or more samples, or quasi-replicated (i.e., significant correlations are found between *similar* predictors and the criterion) in two or more samples. For example, if two different dominance scales are found to be positively correlated with leadership, then we would say that the finding is quasi-replicated. Other quasi-replications are less obvious and depend more upon personal interpretation. Using many similar rather than a few identical predictive variables is a good way to increase the generality of the findings. Both Underwood (1957, pp. 281-287) and Brunswik (1956) have discussed the possible advantages in such a procedure.

Social Achiever

Table 3 shows the correlations between leadership in college and the student's attributes as measured in his senior year of high school.

Table 3 indicates that leadership in college is best predicted by the following high school achievements: Artistic Achievement (2), Number of Elective Offices (4), and Number of Honors Won. The inclusion of oral activities, such as winning debates and public speaking contests, on the high school Artistic Achievement checklist may account for the predictive success of that

[2] The influence of aptitude was not partialed out since the SAT scores have only negligible relationships with achievement in these samples. Correlations range only from −.27 to .19, with 72% falling within ± .10.

TABLE 3

PREDICTION OF SOCIAL ACHIEVEMENT (LEADERSHIP) OVER 1-, 2-, 3-, AND 4-YEAR PERIODS

Variable	Source	Time interval							
		1 year		2 years		3 years		4 years	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1. Leadership	Self-Rating	—	—	—	—	—	—	43c	51c
2. Artistic Achievement (HS)		26	25	35c	41c	20c	18c	20c	
3. Self-Evaluation		14	22	—	—	—	—	29c	33c
4. Number of Elective Offices (HS)		—	—	13	29	—	—	—	—
5. Creative Activities (B)				—	—	—	—		28c
6. Sense of Destiny	DRS			—	—	—	—	23c	28c
7. Popularity	Self-Rating	—	—	18	25	—	—	26c	27c
8. Self-Confidence	Self-Rating	12	27	—	—	—	—	21c	
9. Aggressiveness	Self-Rating							26c	19c
10. Drive to Achieve	Self-Rating	15			25	—	—	18c	27c
11. Number of Honors Won in HS		—	—	—	—	—	—	15	27
12. Dependability	Self-Rating								26c
13. Q_2 Self-Sufficient	16 PF	—	—		−26	—	—	—	—
14. Breadth of Interest		14	25	—	—	—	—		24c
15. Dominance	CPI	—	—	—	—	22	22	—	—
16. Sociability	Self-Rating	—	—	—	—	—	—	22c	
17. Complexity-Simplicity	OPI			—	—	—	—		−21
18. Q_1 Radical	16 PF	—	—		−21	—	—	—	—
19. Self-Understanding	Self—Rating	—	—	—	—	—	—	20c	
20. Creative Activities (A)[a]		13		—	—	—	—	16c	20c
21. Dominance	NMSS	—	—	16	20	—	—	—	—
22. Expressiveness	Self-Rating	—	—	—	—	—	—	19c	
23. Scholarship	Self-Rating			—	—	—	—	16c	19c
24. F Cheerful	16 PF	—	—		19	—	—	—	—
25. Sociability	CPI	—	—	—	—	19	17	—	—
26. M Introverted	16PF	—	—		−19	—	—	—	—
27. Flexibility	CPI	—	—	—	—	−13	−19	—	—
28. Accountant	SVIB	—	—	—	—	—	—		18
29. Barron Originality	OPI	—	—	—	—	—	—	−17	
30. Cheerfulness	Self-Rating			—	—	—	—	12c	17c
31. Super-Ego	NMSS	—	—		17	—	—	—	—
32. Self-Acceptance	CPI	—	—	—	—	15	17	—	—
33. Physicist	SVIB	—	—	—	—	—	—	−16	
34. Originality	Self-Rating		16			—	—	11c	
35. Music (Training Time)		—	—	16		—	—	—	—
36. Sales Manager	SVIB	—	—	—	—	—	—	15	
37. Responsibility	VPI	16		15		—	—		
38. Dependable	Fathers Value		15	—	—	—	—	—	—
39. Inventory of Beliefs (Chicago)		—	—	—	—	−15	−15	—	—
40. Communality	CPI	—	—	—	—	13	15	—	—
41. Perseverance	Self-Rating	14						11c	
42. Achievement via Conformance	CPI	—	—	—	—	14		—	—
43. Independence	Self-Rating	13		—	—	—	—	11c	
44. Verbal Aptitude	SAT					−11	−15		
45. Math Aptitude	SAT						−13		

TABLE 3—Continued

Variable	Source	Time interval: 1 year Boys	1 year Girls	2 years Boys	2 years Girls	3 years Boys	3 years Girls	4 years Boys	4 years Girls
46. Verbal Activity	VPI			13		—	—		
47. Status	VPI			13		—	—	—	—
48. Social Science Teacher	SVIB	—	—	—	—	—	—	13	
49. A Sociable	16 PF	—	—	12		—	—	—	—
50. E Dominance	16 PF	—	—	12		—	—	—	—
51. Socialization	CPI	—	—	—	—	12		—	—
52. Responsibility	CPI	—	—	—	—	12		—	—
53. Capacity for Status	CPI	—	—	—	—	12		—	—
54. Psychological Mindedness	CPI	—	—	—	—		−12	—	—
55. Conformity	VPI			11		—	—		
56. Self-Control	CPI	—	—	—	—		−11	—	—
57. Intellectual Efficiency	CPI	—	—	—	—	10		—	—
58. Achievement via Independence	CPI	—	—	—	—	−09		—	—
59. Good Impression	CPI	—	—	—	—	08		—	—
60. Sense of Well-Being	CPI	—	—	—	—	08		—	—

Note.—Dashes indicate that a given variable was not correlated with the criterion either because it was not available or because it was not used for some other reason. All correlations in this table and succeeding tables are significant at or beyond the .05 level.

• This variable is an a priori scale of creative activities. Subsequently the scores of high and low scorers on this scale were correlated with preference for a different set of activities, school preferences, etc., to produce another Creative Activities scale (Variable Number 5).

variable. Some of the findings with regard to personality scales seem to complement these results: Student leaders have a high Capacity for Status (53), aspire to occupations of high status (47), and resemble men in, or express preferences for, occupations of a social and persuasive nature (Sales Manager, 36; Verbal Activity, 46; Social Science Teacher, 48). In addition, student leaders conceive of themselves as good leaders: concurrent self-ratings of leadership ability have high correlations (.43 for boys and .51 for girls) with leadership in college.

Student leaders conform in many ways to the stereotype of the leader: they are self-confident (3, 6, 8, 32), popular (7), sociable (16, 25, 37, 49, 52), enthusiastic (22, 24, 30, 56), dominant and aggressive (9, 15, 21, 50), responsible (12, 31, 50, 51), conventional (18, 39, 40, 42, 54, 55, 58), and extroverted (26, 54).

There are some inconsistencies in these findings. For example, the student leader's high self-rating on Originality (34) is not supported by the other measures of originality. In fact, the nonleader gets higher scores on originality on certain indices (17, 27, 29). Moreover, the leader's high self-ratings on "independence" are inconsistent with the finding that indicates that he is dependent and lacking in self-sufficiency (13).

Artistic Achiever

Table 4 summarizes the relationships between the student's attributes in high school and his artistic achievement in college.

It is clear from Table 4 that the best predictor of artistic achievement in col-

TABLE 4

PREDICTION OF ARTISTIC ACHIEVEMENT OVER 1-, 2-, 3-, AND 4-YEAR PERIODS

Variable	Source	Time interval							
		1 year		2 years		3 years		4 years	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1. Artistic Achievement (HS)		31	36	35c	41c	32c	38c	28c	24c
2. Expressiveness	Self-Rating			—	—	—	—	30c	28c
3. Originality	Self-Rating	14	18	26	18	—	—	30c	31c
4. Creative Activities (A)		25	21	—	—	—	—	29c	44c
5. Physical Activity (B)	VPI	—	—	—	—	—	—	−25c	−40c
6. Emotionality	VPI	12		29		—	—	16c	
7. Aviator	SVIB	—	—	—	—	—	—	−25	
8. Masculinity-Femininity	SVIB	—	—	—	—	—	—	−24	−18
9. Dominance	CPI	—	—	—	—	22	16	—	—
10. Leadership	Self-Rating			—	—	—	—	21c	21c
11. Physicist	SVIB	—	—	—	—	—	—	−21	−17
12. Conformity	VPI				21	—	—		
13. Status	VPI	14		24		—	—	—	—
14. Dominance	NMSS	—	—	21		—	—	—	—
15. Sales Manager	SVIB	—	—	—	—	—	—	20	18
16. Aggressiveness	Self-Rating			—	—	—	—	18c	20c
17. SAT-Verbal		−27							
18. Accountant	SVIB	—	—	—	—	—	—		−20
19. Self-Confidence	Self-Rating		21	—	—	—	—	15c	
20. Conservatism	Self-Rating		−19	—	—	—	—	−10c	
21. Practical Mindedness	Self-Rating			—	—	—	—		−19c
22. Capacity for Status	CPI	—	—	—	—	19		—	—
23. Scientific Achievement (HS)			18		18				
24. Self-Evaluation		15		—	—	—	—	18c	18c
25. Cheerfulness	Self-Rating	16	18	—	—	—	—		
26. Musician	SVIB	—	—	—	—	—	—		18
27. Sociability	CPI	—	—	—	—	18		—	—
28. High School Grades (HSR)			−18			—	—		
29. Sense of Destiny	DRS			—	—	—	—	17c	17c
30. Social Science Teacher	SVIB	—	—	—	—	—	—	14	17
31. Self-Acceptance	CPI	—	—	—	—	17		—	—
32. Tolerance for Ambiguity	NMSS	—	—	16		—	—	—	—
33. Breadth of Interest			16	—	—	—	—		
34. Complexity-Simplicity	OPI	—	—	—	—	—	—	12	16
35. SAT-Math						−10	−15	−14	
36. College Grades (HPR)							15c		
37. Social Presence	CPI	—	—	—	—	14		—	—
38. Responsibility	VPI			14		—	—		
39. Sociability	Self-Rating			—	—	—	—	13c	
40. Independence	Self-Rating			—	—	—	—	13c	
41. Self-Understanding	Self-Rating			—	—	—	—	13c	
42. M Introverted	16 PF	—	—	13		—	—	—	—
43. E Dominance	16 PF	—	—	13		—	—	—	—
44. Verbal Activity	VPI			13		—	—		
45. Drive to Achieve	Self-Rating			12		—	—	12c	

TABLE 4—Continued

Variable	Source	Time interval							
		1 year		2 years		3 years		4 years	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
46. Athletic Ability	Self-Rating			—	—	—	—	−12c	
47. Intellectuality	VPI	−12				—	—	−12c	
48. Control	VPI			12		—	—	—	—
49. Maturity	Teacher Rating			−12		—	—	—	—
50. Psychological Mindedness	CPI	—	—	—	—		−12	—	—
51. Occupational Level	SVIB	—	—	—	—	—	—	11	
52. Barron Originality	OPI	—	—	—	—	—	—	11	
53. Persistence	NMSS	—	—	11		—	—	—	—
54. Hobby (duration of)		—	—	11		—	—	—	—
55. Creative Activities (B)				—	—	—	—	15c	
56. Self-Control	CPI	—	—	—	—	−09	−11	—	—
57. Psychopathic Deviate	OPI	—	—	—	—	—	—	10	
58. Popularity	Self-Rating			—	—	—	—	10c	
59. Verbal Activity (B)	VPI					—	—	10c	
60. Conformity (B)	VPI					—	—	10c	
61. Communality	CPI	—	—	—	—	−08		—	—

lege is similar achievement in high school. Both concurrent[3] and predictive relationships for high school and collegiate artistic achievement approximate a correlation of .30. High self-ratings on originality and participation in creative activities in high school predict artistic achievement in college with similar efficiency.

The artistic achiever possesses some of the stereotypic traits: originality, tolerance for ambiguity, complexity of outlook, and a preference for creative activities (3, 4, 32, 34, 52, 55). The artistic achiever likes artistic vocations and dislikes occupations in science and the skilled trades (5, 6, 7, 11, 26).

The student artist resembles the student leader in his preference for vocations of a social character (30, 44, 59) and in his high aspirations (13, 51). Like the student leader, he is sociable (27, 37, 38), self-confident (19, 24, 29, 31), expressive (2, 25), and dominant (9, 16, 43), though these qualities are less typical of the student artist than of the student leader.

[3] "Concurrent" refers here to the fact that, in certain samples, the student reported his high school achievements at the same time he reported his college achievements. The similarity between this concurrent correlation and the equivalent predictive one suggests that such retrospective reporting of high school achievements is fairly accurate.

Scientific Achiever

Table 5 summarizes the relationships between the student's attributes in high school and his scientific achievement in college.

Table 5 shows that the best predictor of scientific achievement in college is similar achievement in high school. The remaining predictors which are replicated or quasi-replicated are preferences for scientific and skilled occupations (2, 14, 17), high self and teacher ratings of originality (20, 24), engagement in creative activities (11, 12), and mathematical aptitude (7).

The paucity of findings in Table 5 may be due to the small number of scientific achievements included on the checklist and to the infrequency of such achievements at the college level. On common-sense grounds, we would

TABLE 5

PREDICTION OF SCIENTIFIC ACHIEVEMENT OVER 1-, 2-, 3-, AND 4-YEAR PERIODS

Variable	Source	Time interval							
		1 year		2 years		3 years		4 years	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1. Scientific Achievement (HS)		30	35	21c		20c	25c	16c	24c
2. Physical Activity	VPI			16		—	—	22c	18c
3. Play	NMSS			−15		—	—	—	—
4. Status	VPI			−14		—	—	—	—
5. Artistic Achievement (HS)				−10					
6. Social Leadership	Teacher Rating	—	—	−13		—	—	—	—
7. SAT-Math					12			13	
8. F Cheerful	16 PF	—	—	−11		—	—	—	—
9. College Grades (HPR)						13			
10. Psychological Mindedness	CPI	—	—	—	—	13		—	—
11. Creative Activities (A)		14		—	—	—	—	24c	
12. Creative Activities (B)				—	—	—	—	23c	
13. Scholarship	Self-Rating			—	—	—	—	22c	
14. Physicist	SVIB	—	—	—	—	—	—	18	
15. Drive to Achieve	Self-Rating			—	—	—	—	18c	19c
16. Social Science Teacher	SVIB	—	—	—	—	—	—	−17	
17. Aviator	SVIB	—	—	—	—	—	—	16	
18. Breadth of Interest (HS)				—	—	—	—	14c	
19. Verbal Activity	VPI					—	—	−13c	
20. Originality	Self-Rating		18	—	—	—	—	12c	
21. Emotionality	VPI					—	—	−11c	
22. Self-Confidence	Self-Rating			—	—	—	—	10c	
23. Accountant	SVIB	—	—	—	—	—	—		21
24. Originality	Teacher Rating	12		—	—	—	—		
25. Curious	Fathers Values	12		—	—	—	—	—	—
26. Dependable	Fathers Values	−16		—	—	—	—	—	—

probably not expect to find scientific achievement occurring with much frequency at the undergraduate level.

Academic Achiever

Table 6 summarizes the relationships between the student's attributes in high school and his academic achievement in college.

According to Table 6, the best predictors of academic achievement in college are good grades in high school and high self-ratings of scholarship. Most of the remaining predictors can be arranged in eight clusters which are internally consistent and related psychologically: self-control, persistence, and deferred gratification (5, 14, 21, 27, 32); socialization and super-ego (7, 15, 16, 26); unsociableness and lack of social presence (12, 29, 34); femininity (11, 67); passivity and timidity (37, 71, 74); lack of potential for original behavior (8, 28, 36, 73); restrictive parental attitudes (30, 19, 20, 52, 53, 51, 77); and scholastic aptitudes, verbal and mathematical (23, 24). The academic achiever then can be characterized as a person who got high grades in high

TABLE 6

PREDICTION OF ACADEMIC ACHIEVEMENT

Variable	Source	College grades (HPR)	
		Boys	Girls
1. Scholarship	Self-Rating	29	35
2. Drive to Achieve	Teacher Rating	26	27
3. High School Grades (HSR)		40, 31, 23	40, 36, 20
4. Drive to Achieve	Self-Rating	18	
5. Perseverance	Self-Rating	20	23
6. Self-Evaluation	Self-Rating	18	
7. Citizenship	Teacher Rating	21	20
8. Originality	Teacher Rating	21	22
9. Achievement via Conformance	CPI		28
10. Femininity	CPI	26	19
11. Social Presence	CPI	−26	−19
12. Athletic Ability	Self-Rating		26
13. Deferred Gratification	NMSS	16, 25	
14. Super Ego	NMSS	21	25
15. Responsibility	CPI	16	25
16. Intrusiveness	PARI		23
17. Acceleration of Development	PARI		23
18. Persistence	NMSS	19	23
19. Math Aptitude	SAT	11, 17, 15	15, —
20. Verbal Aptitude	SAT	09, 12	26, —
21. Socialization	CPI	22	22
22. Self-Control	CPI	15	22
23. Complexity-Simplicity			−22
24. Sociability	CPI	−21	
25. Curious	Fathers Values		−17
26. Status	VPI	13	
27. Control	VPI	18, 12	—, 12
28. F Cheerful	16 PF	−20	
29. Capacity for Status	CPI	−19	
30. Q_1 Radical	16 PF		−19
31. E Dominance	16 PF	−13	−19
32. Achievement via Independence	CPI		18
33. G Persistent	16 PF	11	17
34. Cheerfulness	Self-Rating		−17
35. Independence of Judgment			−17
36. Artistic Achievement (HS)			15
37. Intellectuality	VPI	16	
38. Seclusion of Mother	PARI		16
39. Excluding Outside Influences	PARI		16
40. Ascendency of Mother	PARI		16
41. Self-Acceptance	CPI	−15	−15
42. Q_2 Controlled	16 PF		14
43. Introversion	NMSS	08	−14
44. Responsibility	VPI	13	
45. N Shrewd	16 PF	−13	
46. M Introverted	16 PF		−13
47. I Effeminate	16 PF	12	
48. Q_2 Self-Sufficient	16 PF	11	
49. Dependence	NMSS	11	
50. Flexibility	CPI	−10	
51. H Adventurous	16 PF	−08	−12
52. Emotionality	VPI	09	
53. Martyrdom	PARI	09	
54. Dominance	NMSS	08	

school; conceives of himself as a good student; and is self-controlled, persistent, responsible, feminine, and passive. He comes from a family whose restrictiveness and striving attitudes may have been conducive to academic achievement with its emphasis on conformity. While the *criteria* of creative performance are usually unrelated to academic performance (see Table 2), the academic achiever has somewhat less potential for original behavior on *tests* of originality than the student who gets slightly lower grades. This apparent lack of originality may be a natural consequence of his parental background, since certain types of creative performance may be in part a function of impulsive, undersocialized behavior. Both Barron (1953, 1955) and Gough (1953, 1960) have proposed that high grades may be a product of oversocialization and original behavior a product of undersocialization, hypotheses supported by the present findings, which suggest how academic as opposed to creative achievement may occur.

DISCUSSION

Since the findings were obtained with samples of students having very high aptitudes, they cannot be directly applied to students having less scholastic aptitude. The findings reported here are consistent, however, with those reported in previous research, which used similar variables on samples of less able students. For example, the finding that academic achievement is correlated with self-control, persistence, socialization, and super-ego strength reiterates some earlier findings on academic achievement summarized recently by Gowan (1960). However, other clusters of variables associated with academic achievement appear to be opposed to some of the "factors" in Gowan's summary. Gowan reports, for example, that academic achievement is associated with "Anti-authoritarian Behavior"; the present study, which uses subjects more homogeneous with respect to intelligence and socioeconomic status, suggests the opposed hypothesis. It is difficult to make further comparisons between the present findings and Gowan's interpretative summary because of the diversity of the inventories, techniques, and definitions used.

The results for undergraduate achievement in science and art fields appear consistent with several earlier studies. Gough (1953) found that creative performance in these fields is associated with scientific and artistic vocational interests. Taylor, Smith, Ghiselin, and Ellison (1961) report that creative people see themselves as original, and Barron (1953) and MacKinnon (1959) find that such people get high scores on originality measures.

A review of the predictors of student achievement reinforces some axioms about the prediction problem. It is clear that the best single predictor of each of the four kinds of undergraduate achievement is similar achievement in high school, an impressive point when it is realized that the achievement scales (checklists) for both high school and college are extremely brief and have low reliability (Holland, 1961). The limited variability in high school rank and honor point ratio also militates against the efficiency of HSR as a predictor.

The predictive as opposed to the concurrent validity of many of the variables makes explicit the misleading nature of concurrent studies. For example, many of the student variables which had significant concurrent correlations with achievement in high school failed to predict similar achievement in college (cf. Holland, 1961 with the findings in Tables 3-6).

If we compare the predictive effi-

ciency of different types of variables, we see that, next to past achievement, the best predictors are self and teacher ratings, interest inventories, and simple checklists of creative activities and interests. With a few exceptions, True-False type personality and originality scales are next in order. Finally, parental attitude scales and scholastic aptitude tests are the least efficient predictors of later achievement. Taylor et al. (1961), in studies of adult scientists, found that their predictors formed a similar hierarchy of efficiency.

This crude ordering implies that for purposes of prediction we could abandon our efforts to construct sophisticated inventories, and instead, concentrate on securing more elaborate records of past achievement, and more systematic reports of the student's self-conceptions, interests, and daily activities. Personality and originality scales would have to be used concurrently, to give meaning to these new devices and inventories, since biographical and self-reports are frequently difficult to interpret.

The relative success of checklists of past achievement as predictors of collegiate achievement suggests that some of the low or negligible correlations found here could be raised if we revised these brief achievement scales so that they would cover a greater range of achievements. More normal distributions should yield higher correlations.

The intercorrelations of social and artistic achievement suggest that we need to clarify the nature of these achievements and to find ways of distinguishing between them. This problem is part of a larger one—we must find new dimensions of achievement and order these criteria more systematically so that we can study them more thoroughly. The present criteria still represent only a few of the different kinds of achievement.

REFERENCES

BARRON, F. Complexity-simplicity as a personality dimension. *J. abnorm. soc. Psychol.*, 1953, **48**, 163-172.

BARRON, F. The disposition towards originality. *J. abnorm. soc. Psychol.*, 1955, **51**, 478-485.

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

GOUGH, H. G. Some theoretical problems in the construction of practical assessment devices for the early identification of high level talent. Paper read at Social Science Research Council Conference, Princeton, April 1953.

GOUGH, H. G. Theory and measurement of socialization. *J. consult. Psychol.*, 1960, **24**, 23-30.

GOWAN, J. C. Factors of achievement in high school and college. *J. counsel. Psychol.*, 1960, **7**, 91-95.

HOLLAND, J. L. Prediction of scholastic success for a high aptitude sample. *Sch. Soc.*, 1958, **86**, 290-293.

HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *J. educ. Psychol.*, 1959, **50**, 135-142.

HOLLAND, J. L. The prediction of college grades from personality and aptitude variables. *J. educ. Psychol.*, 1960, **51**, 245-254.

HOLLAND, J. L. Creative and academic performance among talented adolescents. *J. educ. Psychol.*, 1961, **52**, 136-147.

MACKINNON, D. W. What do we mean by talent and how do we test for it? In, *The search for talent: College admissions 7.* New York: College Entrance Examination Board, 1960.

TAYLOR, C. W., SMITH, W. R., GHISELIN, B., & ELLISON, R. Explorations in the measurement and prediction of contributions of one sample of scientists. *USAF Aeronaut. Sys. Div. Personnel Lab. tech. Rep.* No. 61-96, 1961.

UNDERWOOD, B. J. *Psychological research.* New York: Appleton-Century-Crofts, 1957.

WEBSTER, H., & HEIST, P. *Manual for the Omnibus Personality Inventory.* Berkeley, Calif.: Center for the Study of Higher Education, 1959. (Mimeo)

(Received December 1, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 3, 144–149

SOME NONINTELLECTUAL CORRELATES OF ACADEMIC ACHIEVEMENT AMONG MEXICAN-AMERICAN SECONDARY SCHOOL STUDENTS

LOIS J. GILL
Denver Public Schools

AND

BERNARD SPILKA
University of Denver

The purpose of this study was to determine personal and maternal correlates of academic achievement among Mexican-American secondary school students. 4 groups (15 each) of achieving and low achieving boys and girls were identified and equated in age, IQ grade level, and courses taken. Employing standard objective measures, it was shown that achievers manifested reliably less hostility and more social maturity, intellectual efficiency, and conformity to rules. Achieving girls and underachieving boys appear to come from strong mother-dominated homes. Since these findings seem meaningful with respect to this subculture, it is suggested that such study of achievement in minority groups may better clarify the nature of relationships among hypothesized variables.

Investigations of achievement (or overachievement) and underachievement have tended to focus upon three main aspects of this problem: (*a*) student personality (McCandless, 1947; McClelland, Atkinson, Clark, & Lowell, 1953; Morgan, 1952), (*b*) parental attitudes relative to scholastic performance (Gough, 1953; Haggard, 1957; Kimball, 1953), and (*c*) socioeconomic correlates of academic achievement (Douvan, 1956; Milner, 1951). Needless to say, research findings relative to the general Anglo-American group oftentimes have been contradictory. Not only have a wide variety of measuring instruments been used, but the group itself has been so heterogeneous that generalizations across regions, classes, school systems, ages, etc. can be made only as very tentative hypotheses for further investigation. The complexity of this problem with the Anglo-American group has been discussed theoretically but has often been ignored in actual study (Gough, 1953; Roth, 1956).

Research relative to possible factors influencing academic achievement among minority groups is lacking even though many of the apparently confounding variables might be more easily controlled and observed with such groups. Cultural determinants usually can be defined with greater accuracy; the socioeconomic distribution of the group is often quite limited; and geographic, rural-urban location, language influences, etc., are easily circumscribed. It was thus the purpose of the present study to investigate relationships between selected nonintellectual student variables and parental attitudes and the academic performance of a group of Mexican-American, lower-class secondary school students. In addition, an effort was made to control academic performance by employing a criterion of achievement on which all subjects were matched.

Literature pertaining to the Mexican-American indicates that this group is characteristically low in academic achievement. There are, however, within this population many students who are academically successful. Jones (1948) and Marden (1952) indicate that the Mexican-American group possesses a culture steeped in custom and tradition with a strongly authoritarian

family pattern of a patriarchal form. Women are highly controlled but are the controlling agent within the home in regard to the children. Predominantly lower class in the United States, Mexican-Americans suffer from the dual frustrations of poverty and prejudice, possibly responding with greater hostility than their lower-class non-Mexican white counterparts (Swickard & Spilka, 1961). Verbal deficiency and other negative influences such as mobility and poor living conditions undoubtedly make school attendance and academic performance a formidable experience.

On the basis of the foregoing the following hypotheses were tested:

1. Assuming that education might be positively valued within this group in the United States, especially where traditional familial controls are also esteemed, it was hypothesized that achieving students would come from homes in which the mother is more dominating, possessive, and ignores the child less.

2. It was hypothesized that achievement of the students coming from this background, would be positively related to: their conformity to the requirements of the academic situation, and the extent of their self-motivation and tendency toward independent action.

3. It also was hypothesized that achievement would be positively related to intellectual efficiency and social maturity.

4. The achieving student, it was further expected, would reveal less hostility and more indications of socially approved and adaptive means of coping with anxiety.

METHOD

Subjects

Sixty high school junior and senior students served as subjects. Thirty had maintained grade point averages (GPA) above the seventieth percentile for their class, while the remaining 30 students had GPAs below the thirtieth percentile. The former group was designated, Achievers, and the latter, Underachievers. The IQs of all subjects were in the 90–110 range, and the groups were matched on IQ, age, sex, and grade. Each group contained 15 boys and 15 girls. Table 1 presents these data. No significant differences were found between sexes, classes, and groups on age or IQ, whereas differences in GPA between groups were as desired, significant well beyond the .01 level.

Tests and Measures

IQ was assessed by the Otis SA Test of Mental Ability (Gamma form). GPAs were based on grades obtained only in academic courses which all students had in common, and which appeared to emphasize the verbal and symbol usage skills measured by the Otis.

Parental variables were evaluated by the Drews and Teahan (1957) modification of the Parent Attitude Survey devised by Shoben (1949). This scale consists of three subscales designed to measure dominating, possessive, and ignoring attitudes of mothers toward their offspring.

The California Psychological Inventory (CPI) (Gough, 1957) was used to assess the following student variables: achievement

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR GROUPS AND SEXES ON THE CONTROL VARIABLES

Variable	Achievers		Underachievers	
	Girls ($N = 15$)	Boys ($N = 15$)	Girls ($N = 15$)	Boys ($N = 15$)
Otis IQ				
M	99.1	98.2	98.9	98.2
SD	6.64	6.04	6.82	6.19
GPA				
M	2.9	2.9	0.7	0.6
SD	1.93	1.69	0.20	0.20
Age in months				
M	207.6	206.7	207.2	207.5
SD	6.39	6.27	6.50	6.46

by conformance (Ac scale), achievement through independent behavior (Ai scale), intellectual efficiency and resourcefulness (Ie scale), and social maturity (So scale).

The Siegel Manifest Hostility Scale (MHS), a 50-item inventory developed by Siegel (1956), was utilized as an objective measure of hostility; and the Jewell Anxiety Adaptation Scale was selected to measure student adaptation to anxiety. Jewell (1959) developed this scale in an attempt to determine differences in individual reaction to anxieties, stressing responses which are performance facilitating in contrast to behavior which interferes with proper performance.

Procedure

The tests were administered to the students during school hours. A code was employed in lieu of names on the forms, and the students were assured that anonymity would be preserved and none of these results would become part of a school record. The mother was chosen to represent parental attitudes. To help insure understanding of the test items, the Drews-Teahan questions were read to each of the mothers. All subjects were highly cooperative.

Statistical evaluation of the obtained data was accomplished through use of a groups-by-sexes analysis of variance design (Lindquist, 1953).

RESULTS

Parental Attitudes

On the Drews-Teahan version of Shoben's Parental Attitude Survey no significant differences were found for the main effects or interaction on the Possessive and Ignoring scales. The Sex × Groups interaction term for the Dominance scale attained significance at the .05 level of confidence ($F = 4.28$, $df = 1/56$). In order to assess this finding the means of groups and sexes were corrected for these main effects and Tukey procedure (Edwards, 1954) was applied. The corrected means were 33.2 for achieving boys and underachieving girls and 35.7 for low achieving boys and high achieving girls. The differences between these mean pairs account for the observed significance at the .05 level of confidence. No significant differences were found between the mothers of the high and low group students on either the possessive or ignoring attitude scales. Hypothesis 1, namely, that achieving students would come from homes in which the mothers tend to be dominating, possessive, and attentive, was partially supported.

Student Variables

There were no significant differences between the responses of achieving and underachieving students on the Ai scale of the CPI. The groups did differ significantly on the three other scales of the CPI. On the Ac scale differences exceeding the .01 level of confidence were found between groups ($F = 75.46$, $df = 1/56$), between sexes ($F = 7.39$, $df = 1/56$), and also for the Sex × Groups interaction term ($F = 33.80$, $df = 1/56$). Significant differences at the .05 level of confidence were also found between the two groups of students on Ie ($F = 5.41$, $df = 1/56$) and So ($F = 7.09$, $df = 1/56$). Hypothesis 2 was supported for achievement via conformance, but not for self motivation and independent action. Hypothesis 3 is also supported in that the achievers appear to be both significantly more intellectually efficient and socially more mature than their underachieving classmates.

The MHS yielded significant differences between the groups in the direction predicted at the .01 level of confidence ($F = 13.22$, $df = 1/56$) and also between sexes ($F = 8.37$, $df = 1/56$). Significant differences, at the .05 level of confidence ($F = 5.73$, $df = 1/56$) were found between groups on Jewell's Anxiety Adaptation Scale. Hypothesis 4, namely, that achieving students would be less hostile and show more socially acceptable means of coping with anxiety, was thus supported.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR GROUPS AND SEXES ON TEST VARIABLES
(N per group = 15)

Variable	Achievers		Underachievers		Significance level		
	Girls	Boys	Girls	Boys	Between groups	Between sexes	Sex & Group interaction
Mother variables:							
Dominance	36.6 (5.67)	31.7 (3.85)	34.5 (6.02)	35.5 (6.82)	*ns*	*ns*	.05
Possessive	24.9 (4.99)	25.1 (7.34)	25.9 (5.09)	22.9 (4.73)	*ns*	*ns*	*ns*
Ignoring	28.1 (4.19)	27.1 (2.35)	27.3 (4.12)	26.3 (3.75)	*ns*	*ns*	*ns*
Student variables:							
Manifest Hostility	15.3 (5.59)	20.4 (6.22)	21.0 (6.76)	25.3 (7.39)	.01	.01	*ns*
Anxiety Adaptation	18.1 (4.70)	18.1 (4.59)	15.1 (6.24)	13.6 (8.15)	.05	*ns*	*ns*
Independent Achievement	17.9 (3.30)	16.3 (4.19)	14.5 (5.08)	15.5 (5.87)	*ns*	*ns*	*ns*
Conformity Achievement	27.5 (3.99)	24.7 (3.91)	21.7 (5.09)	21.8 (5.82)	.01	.01	.01
Intellectual Efficiency	35.1 (5.37)	34.5 (4.75)	31.7 (5.66)	31.5 (5.87)	.05	*ns*	*ns*
Socialization	39.9 (5.65)	36.0 (6.09)	34.7 (5.87)	33.7 (4.60)	.05	*ns*	*ns*

Note.—Standard deviations are given in parentheses.

DISCUSSION

It was hypothesized that achieving students came from more mother-dominated homes than underachievers. The results indicate that no significant differences in mother domination existed between groups but the Sex × Group interaction yielded differences which were significant; that is, underachieving boys and achieving girls come from homes in which the mothers are more dominating than are the mothers of achieving boys and underachieving girls.

In view of this finding, it may be conjectured that restrictive and demanding maternal attitudes are perceived by the academically successful girl as affectional concern for her. The daughter, therefore, can identify with her mother and accept maternal control if the traditional controls are present. The boy, in contrast, may view maternal domination as an infringement on his independence strivings and react negatively to it. In this ethnic group the woman's role traditionally is inferior to the male's position. The male teenager, then, may reject maternal domination and possibly use poor school performance as a weapon against his demanding mother.

The achieving boy who is less domi-

nated may perceive maternal controls as containing elements of reasonableness and affection; he, then, may accept his mother's demands and values and do well in an academic setting. The girl who is less dominated possibly sees this lack of control on the part of her mother as a not-caring attitude and she, in turn, develops an apathetic attitude toward educational attainments.

Mothers of achieving and underachieving students did not differ in their attitudes of possessiveness and ignoring the child. However the high scores on both of these scales in comparison to middle-class, Anglo mothers' responses may indicate that these terms are qualitatively different for Mexican-American families which oftentimes are larger, extremely poor, and with unstable paternal domination.

The achieving student apparently has learned to conform to rules and regulations and to work under supervision. In an academic setting where rapid and diligent work habits are highly valued, the achiever wins good grades. He does not appear to feel hostility toward authority figures, such as teachers, because their demands have served as means for reward and personal satisfaction. He may therefore have developed a strong sense of propriety through an acceptance of rules and customs and thus learned to respond to anxiety produced by educational demand with performance facilitating behavior, by doing the "right" things. Achieving girls apparently conform more than either achieving boys or underachieving girls or boys. This is in keeping with the above finding on maternal dominance, and the role of the female in this ethnic group. The fact that the achieving girls obtained the lowest scores on the MHS further supports this interpretation.

The underachiever who tends to respond to anxiety maladaptively probably is frustrated further by the increasing demands for conforming and academically productive behavior. The scholastically unsuccessful student, who tends to respond on a less efficient intellectual level, has difficulty accepting the demands of authority figures toward whom he feels great hostility. Unable to meet the demands of the school, the underachiever just does not perform scholastically, and thus earns this label.

The lack of significant differences between the groups on the Ai scale of the CPI together with the relatively low scores for all students may be attributed to the fact that independent initiative is not characteristic of low socioeconomic students. Douvan (1956) reports that conditions for the lower class are such that success and failure are less personalized. The group members are familiar with the effects of external factors such as unemployment so there is little need to strive unless success involves some meaningful, material reward. The negative effects of prejudice also could tend to accentuate the highly dependent position of the Mexican-American student.

The generality of these findings, particularly those on the student variables, may be argued. Their meaning however is underscored in the Group × Sexes interaction for maternal dominance. The significance of this variable is especially evident in the Mexican-American culture. It is thus suggested that additional specification of group customs and milieu may similarly contribute to further understanding of nonintellectual factors associated with academic achievement.

REFERENCES

DOUVAN, ELIZABETH. Social status and success strivings. *J. abnorm. soc. Psychol.*, 1956, **52**, 219-223.

DREWS, ELIZABETH N., & TEAHAN, J. E.

Parental attitudes and academic achievement. *J. clin. Psychol.*, 1957, **4**, 328-332.

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

GOUGH, H. G. What determines the academic achievement of high school students. *J. educ. Res.*, 1953, **46**, 321-333.

GOUGH, H. G. *California Psychological Inventory manual.* Palo Alto: Consulting Psychologists Press, 1957.

HAGGARD, E. A. Socialization, personality, and academic achievement in gifted children. *Sch. Rev.*, 1957, **65**, 388-414.

JEWELL, D. S. The effect of anxiety adaptation on performance. Unpublished doctoral dissertation, University of Denver, 1959.

JONES, R. C. Ethnic family patterns: The Mexican family in the United States. *Amer. J. Sociol.*, 1948, **53**, 450-453.

KIMBALL, BARBARA. Case studies in educational failure during adolescence. *Amer. J. Orthopsychiat.*, 1953, **23**, 406-415.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Cambridge, Mass.: Riverside, 1953.

MCCANDLESS, B. The Rorschach as a differential predictor of academic success for matched groups of highly superior men. *Amer. Psychologist*, 1947, **2**, 414-415.

MCCLELLAND, D. C., ATKINSON, J., CLARK, R., & LOWELL, E. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MARDEN, C. F. *Minorities in American society.* New York: American Book, 1952.

MILNER, ESTHER. A study of the relationship between reading readiness in grade one school children and patterns of parent-child interaction. *Child Develpm.*, 1951, **22**, 95-112.

MORGAN, H. H. A psychometric comparison of achieving and nonachieving college students of high ability. *J. consult. Psychol.*, 1952, **16**, 292-298.

ROTH, J. The study of academic success and failure. *Educ. res. Bull.*, 1956, **35**, 176-186.

SHOBEN, E. J. The assessment of parental attitudes in relation to child adjustment. *Genet. psychol. Monogr.*, 1949, **39**, 101-148.

SIEGEL, S. N. The relation of hostility to authoritarianism. *J. abnorm. soc. Psychol.*, 1956, **52**, 368-372.

SWICKARD, D. L., & SPILKA, B. Hostility expression in Spanish-American and non-Spanish white delinquents. *J. consult. Psychol.*, 1961, **25**, 216-220.

(Received December 11, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 3, 150-156

THE EFFECTS OF ANXIETY, ACHIEVEMENT MOTIVATION, AND TASK IMPORTANCE UPON PERFORMANCE ON AN INTELLIGENCE TEST

LAWRENCE S. WRIGHTSMAN, JR.[1]
George Peabody College for Teachers

In a study to determine how anxiety and achievement motivation affect performance on an intelligence test under 2 conditions of task importance, 234 Ss completed the MA scale, a projective measure of achievement motivation, and an intelligence test. 122 Ss were led to believe that performance on the intelligence test was very important; the other 112 Ss were led to believe the test was given only to collect norms. When the test was seen as important, anxious Ss scored significantly lower than nonanxious Ss. In the other condition, anxiety was unrelated to test performance. Ss high in internalized achievement motivation performed equally well on the test regardless of its purported importance. Low achievement motivation Ss did less well when they thought the test was important, perhaps indicating an avoidance reaction to stress.

Although much research has been done recently upon the effects of anxiety and achievement motivation, their influence upon performance in everyday testing situations has not been sufficiently studied. The purpose of this study is to determine how each of these variables influence performance on an intelligence test under different conditions of task importance.

The question of whether a high degree of anxiety has a debilitating effect upon performance on a complex task has been considered by Spence (1958) in a statement of the theory underlying research with the Manifest Anxiety (*MA*) scale (Taylor, 1953). Spence presents evidence for acceptance of an "emotional reactivity hypothesis," which states that "high A subjects have a lower threshold of emotional responsiveness and react with a stronger emotional response than low A subjects to situations containing some degree of stress" (p. 137). A different type of theoretical formulation has been advanced by Sarason and his associates (Ruebush, 1960; Sarason, Davidson, Lighthall, & Waite, 1958), who feel that the highly anxious person acts in a cautious, defensive, dependent way which often impedes his performance on a test.

The stronger emotional response on the part of high anxious subjects in a stressful situation has been found to interfere with their successful completion of a complex task such as is involved in the learning of paired associates (Spence, 1956). Likewise, highly anxious children, who are also found to be more cautious, do less well on a problem solving task than do less anxious, less cautious children (Ruebush, 1960). The complex task selected for investigation in this study was of a different type: performance on an intelligence test. It was assumed that by manipulat-

[1]The author wishes to express his thanks to Carolyn Ashcraft, Ruth Vineberg, and Shirley Wrightsman for their assistance in scoring and data analysis; to John Throne, who administered and scored the *MA* scale data; and to Lorene C. Quay, Norman R. Ellis, and Elizabeth G. French, who made helpful suggestions regarding the interpretation of the data.

ing the importance of the test to the subjects the stress of the situation could also be manipulated. It was hypothesized that when the intelligence test was seen as unimportant there would be no relationship between level of anxiety and score on the intelligence test, but that when the test was seen as important the relationship would be a negative one.

The second variable whose effects were under scrutiny in this study is the level of achievement motivation. In one of the more fruitful approaches to this problem, McClelland and his associates (Clark & McClelland, 1955; Lowell, 1952; McClelland, Atkinson, Clark, & Lowell, 1953) have found that in regard to simple tasks the strength of achievement motivation in some subjects may be experimentally aroused and manipulated through the use of instructions which focus on the importance or unimportance of the task. The question arises as to whether experimental arousal of task importance would be effective in influencing performance on a more complex task such as an intelligence test. The little evidence available on the question of the effects of motivational arousal upon performance in complex tasks is in conflict (Fleishman, 1958; Wrightsman, 1960) and comparisons between studies are difficult to make because of differences in types of subjects and tasks used.

A possible reason for previous negative findings may lie in a failure to consider the level of internalized achievement motivation in the subjects. Achievement motivation theory (McClelland et al., 1953) would predict that the performance of persons high in internalized achievement motivation would be uninfluenced by external conditions such as the purported importance of the task, because their internalized standard of excellence would demand a highly motivated approach to any task. On the other hand, performance of the person low in internalized achievement motivation may more clearly reflect the importance of the task, as it is defined by external conditions. Therefore, it was hypothesized that the purported importance of the task would not influence performance of subjects high in internalized achievement motivation, but would influence the performance of subjects lower in internalized achievement motivation.

METHOD

Normally, all freshmen at George Peabody College for Teachers are administered the American Council on Education Psychological Examination (ACE) and the Cooperative English Tests as a part of the freshman orientation program. After answering these instruments, the 234 freshmen (59 males and 175 females) entering in the fall of 1959 were administered three additional tests. These were the *MA* scale (Taylor, 1953); the Test of Insight, a projective measure of achievement motivation devised by French (1958); and the Adaptability Test (Tiffin & Lawshe, 1954), a timed measure of intelligence. Before taking the last test, 122 subjects were led to believe that the results of this intelligence test would be very important to them, in that the results would follow the student throughout his college career (Important Test condition). The other 112 subjects, members of the Unimportant Test condition, were led to believe that their responses to the intelligence test were being collected for normative purposes only. Subjects in the two conditions were tested at the same time, as the manipulation of the importance of the test was done entirely through the use of printed instructions on the front of the test. After completing the intelligence test, subjects were asked to indicate how important they thought it was for them to do well on the test, by placing a number from 1 through 10 under their name on the test booklet. An analysis indicated that the manipulation was successful in that subjects in the Important Test condition placed significantly greater importance to doing well ($\chi^2 = 5.52$, $p < .02$).

No significant between-condition differences were found on the ACE, the *MA* scale, or the achievement motivation measure. An

estimate of the interscorer reliability of the achievement motivation measure was determined by comparing the experimenter's scoring of the responses of 30 subjects with the scoring done by another person trained in the scoring procedures. The resultant reliability estimate was .83.

There were no significant sex differences on any of the measures. The finding that males and females did not differ in the extent of achievement motivation—a finding in contradiction to some previous research (Pierce & Bowman, 1959)—may be attributed to the use of female names on the form of the Test of Insight administered to female subjects. Despite the absence of sex difference in mean score, the relationships between *MA* scores and measures of aptitude such as the ACE and Adaptability Test are more negative for females than for males. Henceforth, findings for each sex will be reported.

Correlation between the *MA* scale and the achievement motivation measure was −.14. Despite its being significant at the .05 level, the relationship is of such small magnitude that it may be said with confidence that the two scales are measuring essentially different variables.

In both conditions performance on the *MA* scale was significantly negatively correlated with ACE-L and ACE-Q scores (−.19 and −.14 in the Unimportant Test condition and −.28 and −.25 in the Important Test condition). This confirms previous findings that in college samples there is sometimes a slight, but significant, negative relationship between level of anxiety and performance on measures of scholastic aptitude (Spielberger, 1958). Those subjects who were to be in the Important Test condition had higher ACE anxiety correlations than did those who were to be in the Unimportant Test condition. These differences (−.19 vs. −.28 and −.14 vs. −.25) approach but do not reach the .05 level of significance, and there is no clear-cut reason other than a coincidental one for the differences. However, the differences should be borne in mind and the effect of the manipulation should be seen as possibly increasing an already existing near-difference. In an effort to eliminate these differences, partial correlation coefficients were computed in addition to the first-order coefficients.

RESULTS

As is indicated in Table 1, the correlation coefficient between *MA* scores and the Adaptability Test scores in the Unimportant Test condition was a nonsignificant −.06; but in the Important Test condition, the comparable correlation was −.37, significant at the .01 level. The correlations for the two conditions differed significantly at the .01

TABLE 1

CORRELATIONS BETWEEN *MA* SCORE, ADAPTABILITY TEST SCORE, AND ACE SCORES

Correlation	Unimportant Test condition			Important Test condition		
	Males ($N = 32$)	Females ($N = 80$)	Total ($N = 112$)	Males ($N = 27$)	Females ($N = 95$)	Total ($N = 122$)
MA and Adaptibility	.09	−.12	−.06	−.30	−.40**	−.37**
MA and Adaptability, with ACE-L partialed out	.24	.00	.07	−.30	−.24**	−.25**
MA and Adaptability, with ACE-Q partialed out	−.04	.12	.05	−.31**	−.28**	−.28**
MA and ACE-L	−.22	−.18	−.19	−.13	−.32**	−.28**
MA and ACE-Q	.17	−.28**	−.14	−.06	−.31**	−.25**
Adaptability and ACE-L	.50**	.68**	.63**	.71**	.71**	.71**
Adaptability and ACE-Q	.71**	.71**	.71**	.52**	.61**	.59**

** Significant at the .01 level.

level, using a one-tailed test. The relationship was greater in the case of the females than the males, although the difference was not statistically significant.

A similar result occurs when the effects of ACE scores are partialed out; performance on the Adaptability Test is related to *MA* score only under the condition that the test performance is important. It thus appears that when performance on the intelligence test carries little importance, a person's level of anxiety is unrelated to his performance, but when the task is seen as important, persons high in anxiety suffer impaired performance. Sex differences in the Important Test condition are reduced after partialing out the effects of the ACE test.

Further confirmation is found in a comparison of the *MA*-Adaptability correlations in the upper part of Table 1 with the *MA*-ACE correlations in the lower part. Since the ACE was administered to the subjects as a part of the standard freshman testing battery, it may be assumed that, to the Unimportant Test subjects, it bore a level of importance greater than that attached to the Adaptability Test, and to the Important Test subjects, a level of importance *less* than that attached to the Adaptability Test. If this assumption is correct and if anxiety has a debilitating effect on test performance only when the test is considered important, differential effects would be expected in the two conditions. In the Important Test condition, the *MA*-Adaptability correlation should be more negative than the *MA*-ACE correlations, to reflect the greater perceived importance of the Adaptability Test and the consequent deterioration of performance by anxious subjects. For Unimportant Tests subjects, the *MA*-Adaptability correlation should be less negative than the *MA*-ACE correlations, as here performance on the test is presented as unimportant. This differential finding does hold, as is indicated in Table 1, although the differences between the *MA*-Adaptability and *MA*-ACE correlations do not reach a level of statistical significance.

Another indication of effects of anxiety level is revealed in Table 2. Men and women in each condition were di-

TABLE 2

MEAN ADAPTABILITY TEST SCORES FOR SUBJECTS WITH DIFFERENT ANXIETY SCORES
(*N* in cell in parentheses)

Subjects	Unimportant Test condition			Important Test condition		
	Males	Females	Total	Males	Females	Total
Top quarter of *MA* scale Distribution (Scores of 19-36)	21.8 (9)	18.5 (23)	19.4 (32)	15.0 (6)	16.0 (21)	15.8 (27)
Third quarter of *MA* scale Distribution (Scores of 14-18)	18.2 (6)	17.9 (18)	18.0 (24)	19.6 (10)	18.0 (26)	18.5 (36)
Second quarter of *MA* scale Distribution (Scores of 9-13)	20.9 (9)	20.1 (21)	20.4 (30)	23.0 (3)	19.6 (26)	19.9 (29)
First quarter of *MA* scale Distribution (Scores of 1-8)	21.5 (8)	19.5 (18)	20.1 (26)	18.4 (8)	20.5 (22)	19.9 (30)
Totals	20.8 (32)	19.0 (80)	19.6 (112)	18.6 (27)	18.6 (95)	18.6 (122)

vided into four quarters on the basis of *MA* score. Mean Adaptability Test scores for subjects in the Unimportant Test condition bear no relation to the level of anxiety score. In the Important Test condition, it appears that the subjects in the highest quarter of the *MA* distribution are the ones whose Adaptability scores are detrimentally influenced by the stress of the situation. This is true of both sexes, although the number of males in this cell is quite small.

An analysis of variance of the data in Table 2 indicates that mean Adaptability scores did not differ between conditions, but that subjects higher in anxiety did significantly poorer. The findings are in keeping with the results of the correlational analysis presented above.

To study the effects of level of achievement motivation, subjects were divided into high, medium, and low groups on the basis of achievement motivation score. Table 3 presents the results of the analysis for the second hypothesis. The prediction that the importance of the task would influence performance of low achievement motivation subjects but not influence that of high achievement motivation subjects would imply the presence of a significant interaction between treatments and levels of achievement motivation. Table 3 indicates that the interaction was significant at the .01 level. However, neither the importance of the tasks nor the internalized level of achievement motivation in the subjects had any effect upon their scores on the intelligence test.

TABLE 3

ANALYSIS OF VARIANCE TABLE FOR ADAPTABILITY TEST SCORES

Source of Variance	*SS*	*df*	*MS*	*F*
Achievement motivation level	2.52	2	1.26	2.48
Importance treatment	1.36	1	1.36	2.68
Interaction	4.10	2	2.05	4.04**
Error	—	228	0.51	—

Note.—Analysis from Walker and Lev (1953, pp. 381-382).

** Significant at .01 level.

Table 4 presents the mean Adaptability Test scores of subjects with varying levels of achievement motivation. The importance manipulation was successful in influencing Adaptability scores only for the low achievement motivation subjects, who did *poorer* when the test was important than when it was unimportant. This finding occurs in the case of subjects of each sex.

DISCUSSION

The findings regarding the effects of anxiety confirm both the "cautiousness" and the "emotional reactivity" hypotheses and indicate that a stressful situation interferes with the successful performance of high anxious subjects on a complex task. The contribution of this study may lie in the type of complex task used, for the effects upon performance in an intelligence test demonstrated in this study are quite similar to the effects of anxiety arousal on maze learning performance (Sarason, Mandler, & Craighill, 1952) and on the learning of paired associates (Spence, 1956).

If the findings may be generalized to other testing situations, it would seem that the actual test performance of extremely anxious persons is an underestimate of what they can do under more relaxed conditions. The performance of extremely anxious subjects in this study was reduced by almost one standard deviation by the stress of the instructions. The test performance of less anx-

TABLE 4

MEAN ADAPTABILITY TEST SCORES FOR SUBJECTS VARYING IN ACHIEVEMENT MOTIVATION
(*N* in cell in parentheses)

Achievement motivation classification	Unimportant Test condition			Important Test condition		
	Men	Women	Total	Men	Women	Total
High (scores of 5-14)	20.6 (11)	19.2 (23)	19.6 (34)	20.5 (4)	20.4 (30)	20.3 (34)
Medium (scores of 3 and 4)	19.8 (10)	18.9 (34)	19.1 (44)	19.2 (12)	18.7 (38)	18.8 (50)
Low (scores of 0-2)	21.8 (11)	19.1 (23)	20.0 (34)	19.5 (11)	16.4 (27)	16.8 (38)

ious subjects seems unaltered by the stress of the situation.

The findings regarding achievement motivation indicate that subjects with a high drive to do well try equally hard regardless of the task's importance, and corroboration is found for past research indicating that college students perform equally well on an intelligence test whether the test is seen as important or not (Wrightsman, 1960). The attempt to increase level of performance through experimental arousal of motivation was apparently not successful.

The worsened performance of low achievement motivation subjects when the task is important perhaps indicates an almost deliberate failure to exert full effort under the challenge of an important situation. Atkinson (1958, p. 332) states that there seems to be good presumptive evidence that fear of failure is greater in low achievement motivation groups. In the earlier scoring of the Test of Insight, responses indicating both positive and negative aspects of achievement motivation were scored together. The two items of the Test of Insight most likely to elicit responses of fear of failure in achievement were rescored for this type of response alone. (These items, on the female form, were Number 5, "Mary Ann worries a lot about how she has done on tests and examinations" and Number 8, "Janet works harder than most people.") Though much more evidence is necessary, there was very little indication that the low achievement motivation subjects shield themselves from full output on an important task because they fear the outcome.

REFERENCES

ATKINSON, J. W. Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.

CLARK, R. A., & MCCLELLAND, D. C. The effects of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955.

FLEISHMAN, E. A. A relationship between incentive motivation and ability level in psychomotor performance. *J. exp. Psychol.*, 1958, **56**, 78-81.

FRENCH, ELIZABETH G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.

LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31-40.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

PIERCE, J. V., & BOWMAN, P. H. The educational motivation patterns of superior students who do and do not achieve in high school. Washington, D. C.: United States Office of Education (Research Project No. 208: 7136), 1959. (Mimeo)

RUEBUSH, B. K. Interfering and facilitating effects of test anxiety. *J. abnorm. soc. Psychol.*, 1960, **60**, 205-212.

SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., & WAITE, R. R. Rorschach behavior and performance of high and low anxious children. *Child Develpm.*, 1958, **29**, 277-285.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561-565.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven, Conn.: Yale University Press, 1956.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131-141.

SPIELBERGER, C. D. On the relationship between manifest anxiety and intelligence. *J. consult. Psychol.*, 1958, **22**, 220-224.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.

TIFFIN, J., & LAWSHE, C. H. Examiner manual for the Adaptability Test. Chicago: Science Research Associates, 1954.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WRIGHTSMAN, L. S., JR. The effects of the purported validity of a test upon level of motivation and performance. *J. educ. Res.*, 1960, **54**, 153-156.

(Received January 15, 1962)

The Measurement

and Prediction of

TEACHER EFFECTIVENESS

This milestone monograph summarizes 30 years of important research to offer tested guidelines for practitioners and fresh theories for further study.

By

DR. A. S. BARR

and others

156 pages **$5**

(25% discount in lots of 5 or more)

DEMBAR EDUCATIONAL RESEARCH SERVICES

Box 1148 Madison 1, Wisconsin

Deterrents and Reinforcement

THE PSYCHOLOGY OF INSUFFICIENT REWARD

Douglas H. Lawrence and Leon Festinger. Leon Festinger's theory of dissonance is here applied to the problem of why partial reward, delay of reward, and effort expenditure during training result in increased resistance to extinction. The authors present a critical review as well as new data on alternative explanations, and in the final chapter take a look at the broader implications of the new theory. *Stanford Studies in Psychology, II.* $4.75

A Theory of Cognitive Dissonance

Leon Festinger. This book, first published in 1957, has been hailed as one of the seminal contributions to modern psychological theory. "A major theoretical statement on the general topic of communication and social influence."—*Annual Review of Psychology* (1959). "Festinger has developed an idea of considerable power and originality which is well supported by experimental evidence."—*The American Journal of Psychology.* $5.25

Personality and Decision Processes

STUDIES IN THE SOCIAL PSYCHOLOGY OF THINKING

Orville G. Brim, Jr., David C. Glass, David E. Lavin, and Norman Goodman. The development of new concepts and generalizations about the interrelatedness of decision making, personality, and social structure is the focus of these studies, which have been designed and carried out from the point of view of the sociologist. *Stanford Studies in Sociology, II.* $8.00

Order from your bookstore, please

STANFORD UNIVERSITY PRESS